

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |           |  |
|---|-----------|--|
| <b>(51) International Patent Classification <sup>6</sup> :</b><br><b>C07H 21/02, 21/04, 1/00, 14/00, 17/00,</b><br><b>C12Q 1/68, G01N 33/53</b>   | <b>A1</b> | <b>(11) International Publication Number:</b> <b>WO 99/06426</b><br><b>(43) International Publication Date:</b> 11 February 1999 (11.02.99)  |
| <b>(21) International Application Number:</b> PCT/US98/16102<br><b>(22) International Filing Date:</b> 3 August 1998 (03.08.98)<br><br><b>(30) Priority Data:</b><br>60/054,646 4 August 1997 (04.08.97) US<br>60/091,650 2 July 1998 (02.07.98) US<br><br><b>(71) Applicant:</b> MILLENNIUM BIOTHERAPEUTICS, INC.<br>[US/US]; 620 Memorial Drive, Cambridge, MA 02142 (US).<br><br><b>(72) Inventor:</b> PAN, Yang; 6 Hamilton Road #1, Brookline, MA 02146 (US).<br><br><b>(74) Agent:</b> MEIKLEJOHN, Anita, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).  |           | <b>(81) Designated States:</b> AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br><i>With international search report.</i> |
| <b>(54) Title:</b> NOVEL MOLECULES OF THE TANGO-77 RELATED PROTEIN FAMILY AND USES THEREOF<br><br><b>(57) Abstract</b><br><br>Novel Tango-77 polypeptides, proteins, and nucleic acid molecules are disclosed. In addition to isolated, full-length Tango-77 proteins, the invention further provides isolated Tango-77 fusion proteins, antigenic peptides and anti-Tango-77 antibodies. The invention also provides Tango-77 nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecule of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a Tango-77 gene has been introduced or disrupted. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided. |           |  |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |  |    |                          |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   |    |  | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | ML | Mali   | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MN | Mongolia                                     | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MR | Mauritania                                   | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MW | Malawi                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | MX | Mexico                                       | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NE | Niger  | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NL | Netherlands                                  | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NO | Norway                                       | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | NZ | New Zealand                                  |    |                          |
| CM | Cameroon                 |    |  | PL | Poland                                       |    |                          |
| CN | China                    | KR | Republic of Korea                        | PT | Portugal                                     |    |                          |
| CU | Cuba                     | KZ | Kazakstan                                | RO | Romania                                      |    |                          |
| CZ | Czech Republic           | LC | Saint Lucia                              | RU | Russian Federation                           |    |                          |
| DE | Germany                  | LI | Liechtenstein                            | SD | Sudan  |    |                          |
| DK | Denmark                  | LK | Sri Lanka                                | SE | Sweden                                       |    |                          |
| EE | Estonia                  | LR | Liberia                                  | SG | Singapore                                    |    |                          |

- 1 -

NOVEL MOLECULES OF THE TANGO-77 RELATED PROTEIN  
FAMILY AND USES THEREOF

Background of the Invention

The polypeptide cytokine interleukin-1 (IL-1) is a critical mediator of inflammatory and overall immune response. To date, three members of the IL-1 family, IL-1 $\alpha$ , IL-1 $\beta$  and IL-1ra (Interleukin-1 receptor antagonist) have been isolated and cloned. IL-1 $\alpha$  and IL-1 $\beta$  are proinflammatory cytokines which elicit biological responses, whereas IL-1ra is an antagonist of IL-1 $\alpha$  and IL-1 $\beta$  activity. Two distinct cell-surface receptors have been identified for these ligands, the type I IL-1 receptor (IL-1RtI) and type II IL-1 receptor (IL-1RtII). Recent results suggest that the IL-1RtI is the receptor responsible for transducing a signal and producing biological effects.

As mentioned above, IL-1 is a key mediator of the host inflammatory response. While inflammation is an important homeostatic mechanism, aberrant inflammation has the potential for inducing damage to the host. Elevated IL-1 levels are known to be associated with a number of diseases particularly autoimmune diseases and inflammatory disorders.

Since IL-1ra is a naturally occurring inhibitor of IL-1, IL-1ra can be used to limit the aberrant and potentially deleterious effects of IL-1. In experimental animals, pretreatment with IL-1ra has been shown to prevent death resulting from lipopolysaccharide-induced sepsis. The relative absence of IL-1ra has also been suggested to play a role in human inflammatory bowel disease.

Summary of the Invention

The present invention is based, at least in part, on the discovery of a gene encoding Tango-77, a secreted

- 2 -

protein that is predicted to be a member of the cytokine superfamily. The Tango-77 cDNA described below (SEQ ID NO:1) has three possible open reading frames. The first potential open reading frame encompasses 534 nucleotides  
5 extending from nucleotide 356 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:3) and encodes a 178 amino acid protein (SEQ ID NO:2). This protein may include a predicted signal sequence of about 63 amino acids (from about amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID  
10 NO:4) and a predicted mature protein of about 115 amino acids (from about amino acid 64 to amino acid 178 of SEQ ID NO:2 (SEQ ID NO:5)).

The second potential open reading frame encompasses 498 nucleotides extending from nucleotide 389  
15 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:6) and encodes a 167 amino acid protein (SEQ ID NO:7). This protein may include a predicted signal sequence of about 52 amino acids (from about amino acid 1 to about amino acid 52 of SEQ ID NO:7 (SEQ ID NO:8)) and a predicted  
20 mature protein of about 115 amino acids (from about amino acid 52 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:9)).

The third potential open reading frame encompasses 408 nucleotides extending from nucleotide 481 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:10) and encodes  
25 a 136 amino acid protein (SEQ ID NO:11). This protein includes a predicted signal sequence of about 21 amino acids (from about amino acid 1 to about amino acid 21 of SEQ ID NO:11 (SEQ ID NO:12)) and a predicted mature protein of about 115 amino acids (from about amino acid  
30 22 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:13)).

As used herein, the terms "Tango-77", "Tango-77 protein", "Tango-77 polypeptide" and the like, can refer and polypeptide produced by the cDNA of SEQ ID NO:1 including any and all of the Tango-77 gene products  
35 described above.

- 3 -

Tango-77 is expected to inhibit inflammation and play a functional role similar to that of secreted IL-1ra. For example, it is expected that Tango-77 may bind to the IL-1 receptor, thus blocking receptor  
5 activation by inhibiting the binding of IL-1 $\alpha$  and IL-1 $\beta$  to the receptor. Alternatively, Tango-77 may inhibit inflammation through another pathway, for example, by binding to a novel receptor. Accordingly, Tango-77 may be useful as a modulating agent in regulating a variety  
10 of cellular processes including acute and chronic inflammation, e.g., asthma, chronic myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory bowel disease.

In one aspect, the invention provides isolated  
15 nucleic acid molecules encoding Tango-77 or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of Tango-77.

The invention encompasses methods of diagnosing  
20 and treating patients who are suffering from a disorder associated with an abnormal level (undesirably high or undesirably low) of inflammation, abnormal activity of the IL-1 receptor complex, or abnormal activity of IL-1, by administering a compound that modulates the expression  
25 of Tango-77 (at the DNA, mRNA or protein level, e.g., by altering mRNA splicing) or by altering the activity of Tango-77. Examples of such compounds include small molecules, antisense nucleic acid molecules, ribozymes, and polypeptides.

30 The invention features a nucleic acid molecule which is at least 45% (e.g., 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the nucleotide sequence of the cDNA insert of the plasmid

- 4 -

deposited with ATCC as Accession Number (the "cDNA of ATCC 98807"), or a complement thereof.

The invention features a nucleic acid molecule which includes a fragment of at least 100 (e.g., 250,  
5 325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, or 989) nucleotides of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the nucleotide sequence of the cDNA ATCC 98807, or a complement thereof.

10 The invention also features a nucleic acid molecule which includes a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID  
15 NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or the amino acid sequence encoded by the cDNA of ATCC 98807.

In a preferred embodiment, a Tango-77 nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the  
20 nucleotide sequence of the cDNA of ATCC 98807.

Also within the invention is a nucleic acid molecule which encodes a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID  
25 NO:11, SEQ ID NO:12, SEQ ID NO:13, wherein the fragment includes at least 15 (e.g., 25, 30, 50, 100, 150, or 178) contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or the polypeptide  
30 encoded by the cDNA of ATCC Accession Number 98807.

The invention includes a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8,  
35 SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or

- 5 -

an amino acid sequence encoded by the cDNA of ATCC Accession Number 98807, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or a  
5 complement thereof under stringent conditions.

Also within the invention are: an isolated Tango-77 protein having an amino acid sequence that is at least about 45%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:5, SEQ  
10 ID NO:9 or SEQ ID NO:13 (mature human Tango-77), or the amino acid sequence of SEQ ID NO:2, SEQ ID NO:7 or SEQ ID NO:11 (immature human Tango-77).

Also within the invention are: an isolated Tango-77 protein which is encoded by a nucleic acid  
15 molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical to SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the cDNA of ATCC 98807; and an isolated Tango-77 protein which is encoded by a nucleic acid molecule having a nucleotide sequence  
20 which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the non-coding strand of the cDNA of ATCC 98807, or the complement thereof.

25 Also within the invention is a polypeptide which is a naturally occurring allelic variant of a polypeptide that includes the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an  
30 amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID

- 6 -

NO:10 or the complement thereof under stringent conditions.

Another embodiment of the invention features Tango-77 nucleic acid molecules which specifically detect  
5 Tango-77 nucleic acid molecules relative to nucleic acid molecules encoding other members of the cytokine superfamily. For example, in one embodiment, a Tango-77 nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising the  
10 nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement thereof. In another embodiment, the Tango-77 nucleic acid molecule is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, or 989)  
15 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement thereof. In yet another embodiment, the  
20 invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a Tango-77 nucleic acid.

Another aspect of the invention provides a vector, e.g., a recombinant expression vector, comprising a  
25 Tango-77 nucleic acid molecule of the invention. In another embodiment, the invention provides a host cell containing such a vector. The invention also provides a method for producing Tango-77 protein by culturing, in a suitable medium, a host cell of the invention containing  
30 a recombinant expression vector such that a Tango-77 protein is produced.

Another aspect of this invention features isolated or recombinant Tango-77 proteins and polypeptides. Preferred Tango-77 proteins and polypeptides possess at  
35 least one biological activity possessed by naturally

- 7 -

occurring human Tango-77, e.g., (i) the ability to interact with proteins in the Tango-77 signalling pathway (ii) the ability to interact with a Tango-77 ligand or receptor; or (iii) the ability to interact with an  
5 intracellular target protein, (iv) the ability to interact with a protein involved in inflammation and (v) the ability to bind the IL-1 receptor. Other activities include the induction and suppression of polypeptide interleukins, cytokines and growth factors.

10 The Tango-77 proteins of the present invention, or biologically active portions thereof, can be operably linked to a non-Tango-77 polypeptide (e.g., heterologous amino acid sequences) to form Tango-77 fusion proteins. The invention further features antibodies that  
15 specifically bind Tango-77 proteins, such as monoclonal or polyclonal antibodies. In addition, the Tango-77 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

20 In another aspect, the present invention provides a method for detecting the presence of Tango-77 activity or expression in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of Tango-77 activity or expression such that  
25 the presence of Tango-77 activity or expression is detected in the biological sample.

In another aspect, the invention provides a method for modulating Tango-77 activity comprising contacting a cell with an agent that modulates (inhibits or  
30 stimulates)

Tango-77 activity or expression such that Tango-77 activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to Tango-77 protein. In another embodiment, the

- 8 -

agent modulates expression of Tango-77 by modulating transcription of a Tango-77 gene, splicing of a Tango-77 mRNA, or translation of a Tango-77 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the Tango-77 mRNA or the Tango-77 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant Tango-77 protein activity or nucleic acid expression by administering an agent which is a Tango-77 modulator to the subject. In one embodiment, the Tango-77 modulator is a Tango-77 protein. In another embodiment, the Tango-77 modulator is a Tango-77 nucleic acid molecule. In other embodiments, the Tango-77 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant Tango-77 protein or nucleic acid expression can include chronic and acute inflammation.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a Tango-77 protein; (ii) mis-regulation of a gene encoding a Tango-77 protein; and (iii) aberrant post-translational modification of a Tango-77 protein, wherein a wild-type form of the gene encodes a protein with a Tango-77 activity.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a Tango-77 protein. In general, such methods entail measuring a biological activity of a Tango-77 protein in the presence and absence of a test compound and identifying those

- 9 -

compounds which alter the activity of the Tango-77 protein.

The invention also features methods for identifying a compound which modulates the expression of Tango-77 by measuring the expression of Tango-77 in the presence and absence of a compound.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

10                    Brief Description of the Drawings

Figure 1 depicts the cDNA sequence (SEQ ID NO:1) of Tango-77. The Tango-77 cDNA has three possible open reading frames which encode the amino acid sequence (SEQ ID NO:2, SEQ ID NO:7 and SEQ ID NO:11) of human Tango-77. The three potential open reading frames of SEQ ID NO:1 extend from: (1) nucleotide 356 to nucleotide 889 (SEQ ID NO:3); (2) nucleotide 389 to nucleotide 889 (SEQ ID NO:6); and (3) nucleotide 481 to nucleotide 889 (SEQ ID NO:10).

20                    Figure 2 depicts an alignment of an amino acid sequence of Tango-77 (T77; SEQ ID NO:2) with IL-1RA (SEQ ID NO:14), and IL-1 $\beta$  (SEQ ID NO:15).

Figure 3 depicts the genomic sequence of BAC1 (SEQ ID NO:16).

25                    Figure 4 depicts the genomic sequence of BAC2 (SEQ ID NO:17).

Figure 5 depicts an amino acid sequence of an alternatively spliced form of Tango-77 (SEQ ID NO:2) as predicted by Procrustes (T77-procrustes; SEQ ID NO:18).

30                    Figure 6 depicts an alignment of an amino acid sequence of an alternatively spliced form of Tango-77 (T77-procrustes; SEQ ID NO:18) with Tango-77 (SEQ ID NO:2).

- 10 -

Figure 7 depicts an alignment of an amino acid sequence of an alternatively spliced form of Tango-77 (T77-procrustes; SEQ ID NO:18) with IL-1ra (SEQ ID NO:14), and IL-1 $\beta$  (SEQ ID NO:15).

5                    Detailed Description of the Invention

The present invention is based on the discovery of a cDNA molecule encoding human Tango-77, a member of the cytokine superfamily. The cDNA molecule encoding human Tango-77 has three possible open reading frames. The  
10 three possible nucleotide open reading frames for human Tango-77 protein are shown in Figure 1 (SEQ ID NO:3, SEQ ID NO:6 and SEQ ID NO:10). The predicted amino acid sequence for the three possible Tango-77 immature proteins are also shown in  
15 Figure 1 (SEQ ID NO:2, SEQ ID NO:7 or SEQ ID NO:11) and three possible mature proteins are also shown in Figure 1 (SEQ ID NO:5, SEQ ID NO:9 and SEQ ID NO:13).

The Tango-77 cDNA of Figure 1 (SEQ ID NO:1), which is approximately 989 nucleotides long including  
20 untranslated regions, encodes a protein amino acid having a molecular weight of approximately 19 kDa, 18 kDa, or 14.9 kDa (excluding post-translational modifications) and the possible mature form of the protein has a molecular weight of 13 kDa. A plasmid containing a cDNA encoding  
25 human Tango-77 (with the cDNA insert name of Of fthx077) was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia 20110-2209 on July 2, 1998 and assigned Accession Number 98807. This deposit will be maintained under the terms  
30 of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an

- 11 -

admission that a deposit is required under 35 U.S.C. §112.

Human Tango-77 is one member of a family of molecules (the "Tango-77 family") having certain conserved structural and functional features. The term "family," when referring to the protein and nucleic acid molecules of the invention, is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin and a homologue of that protein of murine origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

As used interchangeably herein a "Tango-77 activity", "biological activity of Tango-77" or "functional activity of Tango-77", refers to an activity exerted by a Tango-77 protein, polypeptide or nucleic acid molecule on a Tango-77 responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. A Tango-77 activity can be a direct activity, such as an association with a second protein, or an indirect activity, such as a cellular signaling activity mediated by interaction of the Tango-77 protein with a second protein. In a preferred embodiment, a Tango-77 activity includes at least one or more of the following activities: (i) the ability to interact with proteins in the Tango-77 signalling pathway (ii) the ability to interact with a Tango-77 ligand or receptor; or (iii) the ability to interact with an intracellular target protein, (iv) the ability to interact with a protein involved in

- 12 -

inflammation, and (v) the ability to bind the IL-1 receptor.

Accordingly, another embodiment of the invention features isolated Tango-77 proteins and polypeptides  
5 having a Tango-77 activity.

Yet another embodiment of the invention features Tango-77 molecules which contain a signal sequence. Generally, a signal sequence (or signal peptide) is a peptide containing about 21 to 63 amino acids which  
10 occurs at the extreme N-terminal end of a secretory protein. The native Tango-77 signal sequence (SEQ ID NO:4, SEQ ID NO:8, or SEQ ID NO:12) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells),  
15 expression and/or secretion of Tango-77 can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence. Alternatively, the native Tango-77 signal  
20 sequence can itself be used as a heterologous signal sequence in expression systems, e.g., to facilitate the secretion of a protein of interest.

Various aspects of the invention are described in further detail in the following subsections.

25 I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode Tango-77 proteins or biologically active portions thereof, as well as nucleic acid molecules sufficient for use as hybridization probes  
30 to identify Tango-77-encoding nucleic acids (e.g., Tango-77 mRNA) and fragments for use as PCR primers for the amplification or mutation of Tango-77 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g.,

- 13 -

cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated Tango-77 nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or the complement thereof as a hybridization probe, Tango-77 nucleic acid molecules can be isolated using standard

- 14 -

hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to Tango-77 nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 the cDNA of ATCC 98807, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding Tango-77, for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of Tango-77. The nucleotide sequence determined from the cloning of the human Tango-77 gene allows for the generation of probes and primers designed for use in identifying and/or cloning Tango-77 homologues in other cell types, e.g., from other tissues, as well as Tango-77 homologues from other mammals. The probe/primer typically comprises

- 15 -

substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25,  
5 more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807. Alternatively, the oligonucleotide can typically comprise  
10 a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of a naturally occurring  
15 mutant of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

Probes based on the human Tango-77 nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or identical proteins. The  
20 probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a Tango-77 protein, such as by  
25 measuring a level of a Tango-77-encoding nucleic acid in a sample of cells from a subject, e.g., detecting Tango-77 mRNA levels or determining whether a genomic Tango-77 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically  
30 active portion of Tango-77" can be prepared by isolating a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the nucleotide sequence of the cDNA of ATCC 98807 which encodes a polypeptide having a Tango-77 biological activity, expressing the encoded portion of  
35 Tango-77 protein (e.g., by recombinant expression in

- 16 -

vitro) and assessing the activity of the encoded portion of Tango-77.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807 due to degeneracy of the genetic code and thus encode the same Tango-77 protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

In addition to the human Tango-77 nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of Tango-77 may exist within a population (e.g., the human population). Such genetic polymorphism in the Tango-77 gene may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. As used herein, the term "allelic variant" refers to a nucleotide sequence which occurs at a Tango-77 locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a Tango-77 protein, preferably a mammalian Tango-77 protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the Tango-77 gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations in

- 17 -

Tango-77 that are the result of natural allelic variation and that do not alter the functional activity of Tango-77 are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding Tango-77  
5 proteins from other species (Tango-77 homologues), which have a nucleotide sequence which differs from that of a human Tango-77, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the Tango-77  
10 cDNA of the invention can be isolated based on their identity to the human Tango-77 nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

15 Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, or 989) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule  
20 comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions  
25 for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989),  
30 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at  
35 50-65°C. Preferably, an isolated nucleic acid molecule

- 18 -

of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or the complement thereof, corresponds to a naturally-occurring  
5 nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic  
10 variants of the Tango-77 sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the cDNA of ATCC 98807, thereby  
15 leading to changes in the amino acid sequence of the encoded Tango-77 protein, without altering the biological activity of the Tango-77 protein. Amino acid residues that are not conserved or only semiconserved among Tango-77 of various species may be non-essential for activity  
20 and thus would likely be targets for alteration.

Alternatively, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-  
25 type sequence of Tango-77 (e.g., the sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino  
30 acid residues that are conserved among the Tango-77 proteins of various species may be essential for activity and thus would not likely be targets for alteration, unless one wishes to reduce or alter Tango-77 activity.

Accordingly, another aspect of the invention  
35 pertains to nucleic acid molecules encoding Tango-77

- 19 -

proteins that contain changes in amino acid residues that are not essential for activity. Such Tango-77 proteins differ in amino acid sequence from SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13.

An isolated nucleic acid molecule encoding a Tango-77 protein having a sequence which differs from that of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine,

- 20 -

valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

5 Thus, a predicted nonessential amino acid residue in Tango-77 is preferably replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of a Tango-77 coding sequence, such as by saturation  
10 mutagenesis, and the resultant mutants can be screened for Tango-77 biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

15 In a preferred embodiment, a mutant Tango-77 protein can be assayed for: (1) the ability to form protein:protein interactions with proteins in the Tango-77 signalling pathway; (2) the ability to bind a Tango-77 ligand or receptor; or (3) the ability to bind  
20 to an intracellular target protein or (4) the ability to interact with a protein involved in inflammation or (5) the ability to bind the IL-1 receptor. In yet another preferred embodiment, a mutant Tango-77 can be assayed for the ability to modulate inflammation, asthma,  
25 autoimmune diseases, and sepsis.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-  
30 stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire Tango-77 coding strand, or to only a portion thereof, e.g., all or  
35 part of the protein coding region (or open reading

- 21 -

frame). An antisense nucleic acid molecule can be antisense to a noncoding region of the coding strand of a nucleotide sequence encoding Tango-77. The noncoding regions ("5' and 3' untranslated regions") are the 5' and 5 3' sequences which flank the coding region and are not translated into amino acids.

Given the coding strand sequences encoding Tango-77 disclosed herein (e.g., SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:8), antisense nucleic acids of the invention 10 can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of Tango-77 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or 15 noncoding region of Tango-77 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of Tango-77 mRNA, e.g., an oligonucleotide having the sequence 5'-TGCAACTTTTACAGGAACAC-3' (SEQ ID NO:19) or 20 5'-CCTCACTTTTACCCGAGACTC-3' (SEQ ID NO:20) or 5'-GACGGGTGGTACTTAAACAA-3' (SEQ ID NO:21). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be 25 constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously 30 modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. 35 Examples of modified nucleotides which can be used to

- 22 -

generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a Tango-77 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which

- 23 -

binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

20 An antisense nucleic acid molecule of the invention can be an  $\alpha$ -anomeric nucleic acid molecule. An  $\alpha$ -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual  $\beta$ -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead

- 24 -

ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave Tango-77 mRNA transcripts to thereby inhibit translation of Tango-77 mRNA. A ribozyme having specificity for a Tango-77-encoding nucleic acid can be designed based upon the nucleotide sequence of a Tango-77 cDNA disclosed herein (e.g., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Tango-77-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, Tango-77 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, Tango-77 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the Tango-77 (e.g., the Tango-77 promoter and/or enhancers) to form triple helical structures that prevent transcription of the Tango-77 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) *Bioorganic*

- 25 -

& *Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs of Tango-77 can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of Tango-77 can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996) *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs of Tango-77 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of Tango-77 can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion

- 26 -

would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation  
5 (Hyrup (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) *supra* and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry  
10 and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acid Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to  
15 produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.*  
20 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see,  
25 e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be  
30 modified with hybridization-triggered cleavage agents (see, e.g., Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide,

- 27 -

hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

## II. Isolated Tango-77 Proteins and Anti-Tango-77 Antibodies

5           One aspect of the invention pertains to isolated Tango-77 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-Tango-77 antibodies. In one embodiment, native Tango-77 proteins can be isolated  
10 from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, Tango-77 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a Tango-77 protein or polypeptide  
15 can be synthesized chemically using standard peptide synthesis techniques.

          An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from  
20 the cell or tissue source from which the Tango-77 protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of Tango-77 protein in which the  
25 protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, Tango-77 protein that is substantially free of cellular material includes preparations of Tango-77 protein having less than about 30%, 20%, 10%, or  
30 5% (by dry weight) of non-Tango-77 protein (also referred to herein as a "contaminating protein"). When the Tango-77 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture

- 28 -

medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When Tango-77 protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of Tango-77 protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or non-Tango-77 chemicals.

Biologically active portions of a Tango-77 protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the Tango-77 protein (e.g., the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13), which include fewer amino acids than the full length Tango-77 proteins, and exhibit at least one activity of a Tango-77 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the Tango-77 protein. A biologically active portion of a Tango-77 protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native Tango-77 protein.

Preferred Tango-77 protein has the amino acid sequence shown of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13. Other useful Tango-77 proteins are substantially identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 and retain the functional activity of the protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 yet differ in

- 29 -

amino acid sequence due to natural allelic variation or mutagenesis. Accordingly, a useful Tango-77 protein is a protein which includes an amino acid sequence at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 and retains the functional activity of the Tango-77 proteins of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13. In a preferred embodiment, the Tango-77 protein retains a functional activity of the Tango-77 protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions, e.g., overlapping x 100). Preferably, the two sequences are the same length.

The determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990)

- 30 -

*Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990)

5 *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to Tango-77 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST

10 program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to Tango-77 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402.

15 When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. Another preferred, non-limiting example of a mathematical algorithm utilized for

20 the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a

25 PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating

30 percent identity, only exact matches are counted.

The invention also provides Tango-77 chimeric or fusion proteins. As used herein, a Tango-77 "chimeric protein" or "fusion protein" comprises a Tango-77 polypeptide operably linked to a non-Tango-77

35 polypeptide. A "Tango-77 polypeptide" refers to a

- 31 -

polypeptide having an amino acid sequence corresponding to Tango-77 polypeptides, whereas a "non-Tango-77 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not

5 substantially identical to the Tango-77 protein, e.g., a protein which is different from the Tango-77 protein and which is derived from the same or a different organism. Within a Tango-77 fusion protein the Tango-77 polypeptide can correspond to all or a portion of a Tango-77 protein,

10 preferably at least one biologically active portion of a Tango-77 protein. Within the fusion protein, the term "operably linked" is intended to indicate that the Tango-77 polypeptide and the non-Tango-77 polypeptide are fused in-frame to each other. The non-Tango-77

15 polypeptide can be fused to the N-terminus or C-terminus of the Tango-77 polypeptide.

One useful fusion protein is a GST-Tango-77 fusion protein in which the Tango-77 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins

20 can facilitate the purification of recombinant Tango-77.

In another embodiment, the fusion protein is a Tango-77 protein containing a heterologous signal sequence at its N-terminus. For example, the native Tango-77 signal sequence (i.e., about amino acids 1 to 63

25 of SEQ ID NO:2; SEQ ID NO:4; or about amino acids 1 to 52 of SEQ ID NO:7; SEQ ID NO:8; or about amino acids 1 to 21 of SEQ ID NO:11; SEQ ID NO:12) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression

30 and/or secretion of Tango-77 can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (Ausubel et al., *supra*). Other examples of eukaryotic heterologous

35 signal sequences include the secretory sequences of

- 32 -

melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the *phoA* secretory signal (Sambrook et al.,  
5 *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an Tango-77-immunoglobulin fusion protein in which all or part of Tango-77 is fused to sequences derived from a  
10 member of the immunoglobulin protein family. The Tango-77-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a Tango-77 ligand and a Tango-77 receptor on the  
15 surface of a cell, to thereby suppress Tango-77-mediated signal transduction *in vivo*. The Tango-77-immunoglobulin fusion proteins can be used to affect the bioavailability of a Tango-77 cognate ligand. Inhibition of the Tango-77 ligand/Tango-77 interaction may be useful therapeutically  
20 for both the treatment of inflammatory and autoimmune disorders. Moreover, the Tango-77-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-Tango-77 antibodies in a subject, to purify Tango-77 ligands and in screening assays to identify  
25 molecules which inhibit the interaction of Tango-77 with a Tango-77 receptor.

Preferably, a Tango-77 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the  
30 different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as  
35 appropriate, alkaline phosphatase treatment to avoid

- 33 -

undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, e.g., *Current Protocols in Molecular Biology*, Ausubel et al. eds., John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An Tango-77-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the Tango-77 protein.

The present invention also pertains to variants of the Tango-77 proteins (i.e., proteins having a sequence which differs from that of the Tango-77 amino acid sequence). Such variants can function as either Tango-77 agonists (mimetics) or as Tango-77 antagonists. Variants of the Tango-77 protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the Tango-77 protein. An agonist of the Tango-77 protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the Tango-77 protein. An antagonist of the Tango-77 protein can inhibit one or more of the activities of the naturally occurring form of the Tango-77 protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the Tango-77 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer

- 34 -

side effects in a subject relative to treatment with the naturally occurring form of the Tango-77 proteins.

Variants of the Tango-77 protein which function as either Tango-77 agonists (mimetics) or as Tango-77 antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the Tango-77 protein for Tango-77 protein agonist or antagonist activity. In one embodiment, a variegated library of Tango-77 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of Tango-77 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential Tango-77 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of Tango-77 sequences therein. There are a variety of methods which can be used to produce libraries of potential Tango-77 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential Tango-77 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the Tango-77 protein coding sequence can be used to generate a variegated population of Tango-77 fragments for

- 35 -

screening and subsequent selection of variants of a Tango-77 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a Tango-77 coding sequence with  
5 a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed  
10 duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the Tango-77 protein.

15 Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the  
20 gene libraries generated by the combinatorial mutagenesis of Tango-77 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors,  
25 transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble  
30 mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify Tango-77 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993)  
35 *Protein Engineering* 6(3):327-331).

- 36 -

An isolated Tango-77 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind Tango-77 using standard techniques for polyclonal and monoclonal antibody preparation. The  
5 full-length Tango-77 protein can be used or, alternatively, the invention provides antigenic peptide fragments of Tango-77 for use as immunogens. The antigenic peptide of Tango-77 comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the  
10 amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13 and encompasses an epitope of Tango-77 such that an antibody raised against the peptide forms a specific immune complex with Tango-77.

15 A Tango-77 immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed Tango-77 protein or a  
20 chemically synthesized Tango-77 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic Tango-77 preparation induces  
25 a polyclonal anti-Tango-77 antibody response.

Accordingly, another aspect of the invention pertains to anti-Tango-77 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of  
30 immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as Tango-77. A molecule which specifically binds to Tango-77 is a molecule which binds Tango-77, but does not substantially bind other molecules in a sample, e.g., a  
35 biological sample, which naturally contains Tango-77.

- 37 -

Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides  
5 polyclonal and monoclonal antibodies that bind Tango-77. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a  
10 particular epitope of Tango-77. A monoclonal antibody composition thus typically displays a single binding affinity for a particular Tango-77 protein with which it immunoreacts.

Polyclonal anti-Tango-77 antibodies can be  
15 prepared as described above by immunizing a suitable subject with a Tango-77 immunogen. The anti-Tango-77 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized  
20 Tango-77. If desired, the antibody molecules directed against Tango-77 can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after  
25 immunization, e.g., when the anti-Tango-77 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein  
30 (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for  
35 producing hybridomas is well known (see generally Current

- 38 -

Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a  
5 Tango-77 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds Tango-77.

Any of the many well known protocols used for  
10 fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-Tango-77 monoclonal antibody (see, e.g., Current Protocols in Immunology, *supra*; Galfre et al. (1977) *Nature* 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension*  
15 *In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) *Yale J. Biol. Med.*, 54:387-402. Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the  
20 immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized  
25 mouse cell line, e.g., a myeloma cell line that is sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-  
30 NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using  
35 HAT medium, which kills unfused and unproductively fused

- 39 -

myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind Tango-77, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-Tango-77 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with Tango-77 to thereby isolate immunoglobulin library members that bind Tango-77. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734.

Additionally, recombinant anti-Tango-77 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art,

- 40 -

for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; 5 U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura 10 et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 15 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice 20 which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of Tango-77. 25 Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic 30 mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion 35 of this technology for producing human antibodies and

- 41 -

human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition,  
5 companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to the described above.

Completely human antibodies which recognize a  
10 selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope.

15 First, a non-human monoclonal antibody which binds a selected antigen (epitope), e.g., an antibody which inhibits Tango-77 activity, is identified. The heavy chain and the light chain of the non-human antibody are cloned and used to create phage display Fab fragments.  
20 For example, the heavy chain gene can be cloned into a plasmid vector so that the heavy chain can be secreted from bacteria. The light chain gene can be cloned into a phage coat protein gene so that the light chain can be expressed on the surface of phage. A repertoire (random  
25 collection) of human light chains fused to phage is used to infect the bacteria which express the non-human heavy chain. The resulting progeny phage display hybrid antibodies (human light chain/non-human heavy chain). The selected antigen is used in a panning screen to  
30 select phage which bind the selected antigen. Several rounds of selection may be required to identify such phage. Next, human light chain genes are isolated from the selected phage which bind the selected antigen. These selected human light chain genes are then used to  
35 guide the selection of human heavy chain genes as

- 42 -

follows. The selected human light chain genes are inserted into vectors for expression by bacteria. Bacteria expressing the selected human light chains are infected with a repertoire of human heavy chains fused to phage. The resulting progeny phage display human antibodies (human light chain/human heavy chain).

Next, the selected antigen is used in a panning screen to select phage which bind the selected antigen. The phage selected in this step display completely human antibody which recognize the same epitope recognized by the original selected, non-human monoclonal antibody.... The genes encoding both the heavy and light chains are readily isolated and be further manipulated for production of human antibody. This technology is described by Jespers et al. (1994, *Bio/technology* 12:899-903).

An anti-Tango-77 antibody (e.g., monoclonal antibody) can be used to isolate Tango-77 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-Tango-77 antibody can facilitate the purification of natural Tango-77 from cells and of recombinantly produced Tango-77 expressed in host cells. Moreover, an anti-Tango-77 antibody can be used to detect Tango-77 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the Tango-77 protein. Anti-Tango-77 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials.

- 43 -

Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and  
5 avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of  
10 bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

### III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to  
15 vectors, preferably expression vectors, containing a nucleic acid molecule encoding Tango-77 (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of  
20 vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous  
25 replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon  
30 introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant

- 44 -

DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression

- 45 -

vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., Tango-77 proteins, mutant forms  
5 of Tango-77, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of Tango-77 in prokaryotic or eukaryotic cells, e.g., bacterial cells such as *E. coli*, insect cells (using baculovirus  
10 expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and  
15 translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the  
20 expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant  
25 protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction  
30 of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase.  
35 Typical fusion expression vectors include pGEX (Pharmacia

- 46 -

Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al. (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident  $\lambda$  prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the Tango-77 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1

- 47 -

(Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp,  
5 San Diego, CA).

Alternatively, Tango-77 can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series  
10 (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a  
15 mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral  
20 regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al. (*supra*).

25 In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory  
30 elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular  
35 promoters of T cell receptors (Winoto and Baltimore

- 48 -

(1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine *hox* promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the  $\alpha$ -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to Tango-77 mRNA. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see

- 49 -

Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, Tango-77 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome.

- 50 -

In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable  
5 markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding Tango-77 or can be introduced on a separate vector. Cells stably  
10 transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a  
15 prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) Tango-77 protein. Accordingly, the invention further provides methods for producing Tango-77 protein using the host cells of the invention. In one embodiment, the method comprises  
20 culturing the host cell of invention (into which a recombinant expression vector encoding Tango-77 has been introduced) in a suitable medium such that Tango-77 protein is produced. In another embodiment, the method further comprises isolating Tango-77 from the medium or  
25 the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which  
30 Tango-77-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous Tango-77 sequences have been introduced into their genome or homologous recombinant animals in which endogenous Tango-77  
35 sequences have been altered. Such animals are useful for

- 51 -

studying the function and/or activity of Tango-77 and for identifying and/or evaluating modulators of Tango-77 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous Tango-77 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing Tango-77-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The Tango-77 cDNA sequence e.g., that of (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6; SEQ ID NO:10 or the cDNA of ATCC 98807) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human Tango-77 gene, such as a mouse Tango-77 gene, can be isolated based on hybridization to the human Tango-77 cDNA and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency

- 52 -

of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the Tango-77 transgene to direct expression of Tango-77 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the Tango-77 transgene in its genome and/or expression of Tango-77 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding Tango-77 can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a Tango-77 gene (e.g., a human or a non-human homolog of the Tango-77 gene, e.g., a murine Tango-77 gene) into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the Tango-77 gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous Tango-77 gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous Tango-77 gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby

- 53 -

alter the expression of the endogenous Tango-77 protein). In the homologous recombination vector, the altered portion of the Tango-77 gene is flanked at its 5' and 3' ends by additional nucleic acid of the Tango-77 gene to  
5 allow for homologous recombination to occur between the exogenous Tango-77 gene carried by the vector and an endogenous Tango-77 gene in an embryonic stem cell. The additional flanking Tango-77 nucleic acid is of sufficient length for successful homologous recombination  
10 with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic  
15 stem cell line (e.g., by electroporation) and cells in which the introduced Tango-77 gene has homologously recombined with the endogenous Tango-77 gene are selected (see, e.g., Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal  
20 (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and  
25 the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing  
30 homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

- 54 -

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

- 55 -

IV. Pharmaceutical Compositions

The Tango-77 nucleic acid molecules, Tango-77 proteins, and anti-Tango-77 antibodies (also referred to herein as "active compounds") of the invention can be  
5 incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is  
10 intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active  
15 substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

20 A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, (e.g. intravenous, intradermal, subcutaneous) (e.g., oral inhalation), transdermal  
25 (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene  
30 glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as  
35 acetates, citrates or phosphates and agents for the

- 56 -

adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable  
5 syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable  
10 solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be  
15 fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing,  
20 for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance  
25 of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid,  
30 thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including

- 57 -

in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a Tango-77  
5 protein or anti-Tango-77 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a  
10 sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and  
15 freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in  
20 gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier  
25 for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and  
30 the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a  
35 lubricant such as magnesium stearate or Sterotes; a

- 58 -

glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

5 For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

10 Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and  
15 include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are  
20 formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention  
25 enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and  
30 microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to  
35 those skilled in the art. The materials can also be

- 59 -

obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as  
5 pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or  
10 parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active  
15 compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the  
20 particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors.  
25 Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, e.g., Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the  
30 gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g.  
35 retroviral vectors, the pharmaceutical preparation can

- 60 -

include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with  
5 instructions for administration.

#### V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening  
10 assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). A  
15 Tango-77 protein interacts with other cellular proteins and can thus be used for regulation of inflammation. The polypeptides of the invention can be used in assays to determine biological activity. For example, they could be used in a panel of proteins for high-throughput  
20 screening.

The isolated nucleic acid molecules of the invention can be used to express Tango-77 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect Tango-77 mRNA  
25 (e.g., in a biological sample) or a genetic lesion in a Tango-77 gene, and to modulate Tango-77 activity. In addition, the Tango-77 proteins can be used to screen drugs or compounds which modulate the Tango-77 activity or expression as well as to treat disorders characterized  
30 by insufficient or excessive production of Tango-77 protein or production of Tango-77 protein forms which have decreased or aberrant activity compared to Tango-77 wild type protein. In addition, the anti-Tango-77

- 61 -

antibodies of the invention can be used to detect and isolate Tango-77 proteins and modulate Tango-77 activity.

This invention further pertains to novel agents identified by the above-described screening assays and  
5 uses thereof for treatments as described herein.

#### A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents  
10 (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to Tango-77 proteins or have a stimulatory or inhibitory effect on, for example, Tango-77 expression or Tango-77 activity.

Examples of methods for the synthesis of molecular  
15 libraries can be found in the art, for example in:

DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909;  
Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422;  
Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.  
20

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-  
25 421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent Nos. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith  
30 (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

- 62 -

In another embodiment, an assay is used to determine the ability of the test compound to modulate the activity of Tango-77 or a biologically active portion thereof, for example, by determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule. As used herein, a "target molecule" is a molecule with which a Tango-77 protein binds or interacts in nature, for example, a molecule on the surface of a cell. A Tango-77 target molecule can be a non-Tango-77 molecule or a Tango-77 protein or polypeptide of the present invention. In one embodiment, a Tango-77 target molecule is a component of a signal transduction pathway, for example, Tango-77 may bind to a IL-1 receptor or another receptor thereby blocking the receptor and inhibiting future signal transduction. Determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule can be accomplished by one of the methods described above. In a preferred embodiment, determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular  $\text{Ca}^{2+}$ , diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a Tango-77-responsive regulatory element operably linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, inflammation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a Tango-77 protein or biologically active portion thereof with a test compound and determining the ability of the

- 63 -

test compound to bind to the Tango-77 protein or biologically active portion thereof. Binding of the test compound to the Tango-77 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the Tango-77 protein or biologically active portion thereof with a known compound which binds Tango-77 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a Tango-77 protein, wherein determining the ability of the test compound to interact with a Tango-77 protein comprises determining the ability of the test compound to preferentially bind to Tango-77 or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting Tango-77 protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the Tango-77 protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of Tango-77 can be accomplished, for example, by determining the ability of the Tango-77 protein to bind to a Tango-77 target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of Tango-77 can be accomplished by determining the ability of the Tango-77 protein to further modulate a Tango-77 target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the Tango-77 protein or biologically

- 64 -

active portion thereof with a known compound which binds Tango-77 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a Tango-77 protein, wherein determining the ability of the test compound to interact with a Tango-77 protein comprises determining the ability of the Tango-77 protein to preferentially bind to or modulate the activity of a Tango-77 target molecule.

10 It is possible that membrane-bound forms of Tango-77 exist. The cell-free assays of the present invention are amenable to use of both the forms Tango-77. In the case of cell-free assays comprising a membrane-bound form of Tango-77, it may be desirable to utilize a  
15 solubilizing agent such that the membrane-bound form of Tango-77 is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide,  
20 Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.  
25

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either Tango-77 or its target molecule to facilitate separation of complexed from uncomplexed forms  
30 of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to Tango-77, or interaction of Tango-77 with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for  
35 containing the reactants. Examples of such vessels

- 65 -

include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For  
5 example, glutathione-S-transferase/ Tango-77 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical Co.; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined  
10 with the test compound or the test compound and either the non-adsorbed target protein or Tango-77 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or  
15 microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of Tango-77 binding or activity  
20 determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either Tango-77 or its target molecule can be immobilized utilizing conjugation of  
25 biotin and streptavidin. Biotinylated Tango-77 or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated  
30 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with Tango-77 or target molecules but which do not interfere with binding of the Tango-77 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or Tango-77  
35 trapped in the wells by antibody conjugation. Methods

- 66 -

for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the Tango-77 or target molecule, as well as  
5 enzyme-linked assays which rely on detecting an enzymatic activity associated with the Tango-77 or target molecule.

In another embodiment, modulators of Tango-77 expression are identified in a method in which a cell is contacted with a candidate compound and the expression of  
10 Tango-77 mRNA or protein in the cell is determined. The level of expression of Tango-77 mRNA or protein in the presence of the candidate compound is compared to the level of expression of Tango-77 mRNA or protein in the absence of the candidate compound. The candidate  
15 compound can then be identified as a modulator of Tango-77 expression based on this comparison. For example, when expression of Tango-77 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence,  
20 the candidate compound is identified as a stimulator of Tango-77 mRNA or protein expression. Alternatively, when expression of Tango-77 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate  
25 compound is identified as an inhibitor of Tango-77 mRNA or protein expression. The level of Tango-77 mRNA or protein expression in the cells can be determined by methods described herein for detecting Tango-77 mRNA or protein.

30 In yet another aspect of the invention, the Tango-77 proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054;  
35 Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi

- 67 -

et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with Tango-77 ("Tango-77-binding proteins" or "Tango-77-bp") and modulate Tango-77 activity. Such  
5 Tango-77-binding proteins are also likely to be involved in the propagation of signals by the Tango-77 proteins as, for example, upstream or downstream elements of the Tango-77 pathway.

The two-hybrid system is based on the modular  
10 nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for Tango-77 is fused to a gene encoding the DNA binding domain of a known  
15 transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If  
20 the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an Tango-77-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ)  
25 which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes  
30 the protein which interacts with Tango-77.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

- 68 -

### B. Detection Assays

Portions or fragments of the cDNA sequence identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, the sequence can be used to: (i) map the respective gene on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

#### 1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome.

Accordingly, Tango-77 nucleic acid molecules described herein or fragments thereof, can be used to map the location of the Tango-77 gene(s) on a chromosome. The mapping of the Tango-77 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, a Tango-77 gene can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the Tango-77 sequences. Computer analysis of Tango-77 sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the Tango-77 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and

- 69 -

mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow (because they lack a particular enzyme) but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the Tango-77 sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a Tango-77 sequence to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical, e.g., colcemid that disrupts the mitotic spindle. The

- 70 -

chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be  
5 used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases  
10 will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al. (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used  
15 individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding  
20 sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the  
25 sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the  
30 same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) Nature 325:783-787.

Moreover, differences in the DNA sequences between  
35 individuals affected and unaffected with a disease

- 71 -

associated with the Tango-77 gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and  
5 unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that  
10 DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

## 2. Tissue Typing

15 The Tango-77 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its  
20 personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost,  
25 switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present  
30 invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the Tango-77 sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the

- 72 -

sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The Tango-77 sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from Tango-77 sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

- 73 -

### 3. Use of Partial Tango-77 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the Tango-77 sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 or 30 bases.

The Tango-77 sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for

- 74 -

example, an in situ hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such Tango-77 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., Tango-77 primers or probes can be used to screen tissue culture for contamination (i.e., screen for the presence of a mixture of different types of cells in a culture).

### C. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining Tango-77 protein and/or nucleic acid expression as well as Tango-77 activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant Tango-77 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with Tango-77 protein, nucleic acid expression or activity. For example, mutations in a Tango-77 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with Tango-77 protein, nucleic acid expression or activity.

- 75 -

Another aspect of the invention provides methods for determining Tango-77 protein, nucleic acid expression or Tango-77 activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of Tango-77 in clinical trials.

These and other agents are described in further detail in the following sections.

#### 1. Diagnostic Assays

An exemplary method for detecting the presence or absence of Tango-77 in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting Tango-77 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes Tango-77 protein such that the presence of Tango-77 is detected in the biological sample. A preferred agent for detecting Tango-77 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to Tango-77 mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length Tango-77 nucleic acid, such as the nucleic acid of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent

- 76 -

conditions to Tango-77 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting Tango-77 protein is an antibody capable of binding to Tango-77 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')<sub>2</sub>) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect Tango-77 mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of Tango-77 mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of Tango-77 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of Tango-77 genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of Tango-77 protein include introducing into a subject a labeled anti-Tango-77 antibody. For example, the antibody can be

- 77 -

labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains  
5 protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means  
10 from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting Tango-77 protein, mRNA, or  
15 genomic DNA, such that the presence of Tango-77 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of Tango-77 protein, mRNA or genomic DNA in the control sample with the presence of Tango-77 protein, mRNA or genomic DNA in the test sample.

20 The invention also encompasses kits for detecting the presence of Tango-77 in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of Tango-77  
25 (e.g., an immunological disorder). For example, the kit can comprise a labeled compound or agent capable of detecting Tango-77 protein or mRNA in a biological sample and means for determining the amount of Tango-77 in the sample (e.g., an anti-Tango-77 antibody or an  
30 oligonucleotide probe which binds to DNA encoding Tango-77, e.g., SEQ ID NO:1 or SEQ ID NO:3 or SEQ ID NO:6, or SEQ ID NO:10). Kits may also include instruction for observing that the tested subject is suffering from or is at risk of developing a disorder  
35 associated with aberrant expression of Tango-77 if the

- 78 -

amount of Tango-77 protein or mRNA is above or below a normal level.

For antibody-based kits, the kit may comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to Tango-77 protein; and, optionally (2) a second, different antibody which binds to Tango-77 protein or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit may comprise, for example: (1) an oligonucleotide, e.g., a detectably labelled oligonucleotide, which hybridizes to a Tango-77 nucleic acid sequence or (2) a pair of primers useful for amplifying a Tango-77 nucleic acid molecule;

The kit may also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit may also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit may also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of Tango-77.

## 2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant Tango-77 expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or

- 79 -

at risk of developing a disorder associated with aberrant expression or activity. Thus, the present invention provides a method in which a test sample is obtained from a subject and Tango-77 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of Tango-77 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant Tango-77 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant Tango-77 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease Tango-77 activity). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant Tango-77 expression or activity in which a test sample is obtained and Tango-77 protein or nucleic acid is detected (e.g., wherein the presence of Tango-77 protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant Tango-77 expression or activity).

The methods of the invention can also be used to detect genetic lesions or mutations in a Tango-77 gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant

- 80 -

inflammation. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting  
5 the integrity of a gene encoding a Tango-77-protein, or the mis-expression of the Tango-77 gene. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from a Tango-77 gene;  
10 2) an addition of one or more nucleotides to a Tango-77 gene; 3) a substitution of one or more nucleotides of a Tango-77 gene; 4) a chromosomal rearrangement of a Tango-77 gene; 5) an alteration in the level of a messenger RNA transcript of a Tango-77 gene; 6) an  
15 aberrant modification of a Tango-77 gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a Tango-77 gene; 8) a non-wild type level of a Tango-77-protein; 9) an allelic loss of a Tango-77  
20 gene, and 10) an inappropriate post-translational modification of a Tango-77-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions or mutations in a Tango-77 gene. A preferred biological sample is a  
25 peripheral blood leukocyte sample isolated by conventional means from a subject.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and  
30 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for  
35 detecting point mutations in the Tango-77-gene (see,

- 81 -

e.g., Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a Tango-77 gene under conditions such that hybridization and amplification of the Tango-77-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a Tango-77 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA

- 82 -

indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in Tango-77 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotide probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in Tango-77 can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al. supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the Tango-77 gene and detect mutations by comparing the sequence of the sample Tango-77 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a

- 83 -

variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT Publication No. WO 94/16101; 5 Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the Tango-77 gene include methods in which protection from 10 cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the 15 wild-type Tango-77 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and 20 sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium 25 tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the 30 control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize 35 mismatched base pairs in double-stranded DNA (so called

- 84 -

"DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in Tango-77 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the  
5 thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a Tango-77 sequence, e.g., a wild-type Tango-77 sequence, is hybridized to a cDNA or other DNA product  
10 from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in  
15 electrophoretic mobility will be used to identify mutations in Tango-77 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc.*  
20 *Natl. Acad. Sci. USA* 86:2766; see also Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control Tango-77 nucleic acids will be denatured and allowed to renature. The secondary structure of single-  
25 stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by  
30 using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in

- 85 -

electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing  
5 a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of  
10 approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

15 Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the  
20 known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides  
25 are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification  
30 technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on  
35 differential hybridization) (Gibbs et al. (1989) *Nucleic*

- 86 -

Acids Res. 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to  
5 introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany  
10 (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of  
15 amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used,  
20 e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a Tango-77 gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which Tango-77 is  
25 expressed may be utilized in the prognostic assays described herein.

### 3. Pharmacogenomics

Agents, or modulators which have a stimulatory or  
30 inhibitory effect on Tango-77 activity (e.g., Tango-77 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., acute or chronic inflammation and asthma)  
35 associated with aberrant Tango-77 activity. In

- 87 -

conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens.

Accordingly, the activity of Tango-77 protein, expression of Tango-77 nucleic acid, or mutation content of Tango-77 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-

- 88 -

malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM shows no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of Tango-77 protein, expression of Tango-77 nucleic acid, or mutation content of Tango-77 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes

- 89 -

to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a Tango-77 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

#### 4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of Tango-77 (e.g., the ability to modulate aberrant inflammation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase Tango-77 gene expression, increase protein levels, or upregulate Tango-77 activity, can be monitored in clinical trials of subjects exhibiting decreased Tango-77 gene expression, decreased protein levels, or downregulated Tango-77 activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease Tango-77 gene expression, decrease protein levels, or downregulate Tango-77 activity, can be monitored in clinical trials of subjects exhibiting increased Tango-77 gene expression, increased protein levels, or upregulated Tango-77 activity.

For example, and not by way of limitation, genes, including Tango-77, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates Tango-77 activity (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared

- 90 -

and analyzed for the levels of expression of Tango-77 and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as  
5 described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of Tango-77 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of  
10 the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention  
15 provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the  
20 steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a Tango-77 protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples  
25 from the subject; (iv) detecting the level of expression or activity of the Tango-77 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the Tango-77 protein, mRNA, or genomic DNA in the pre-administration sample  
30 with the Tango-77 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or  
35 activity of Tango-77 to higher levels than detected,

- 91 -

i.e., to increase the effectiveness of the agent.  
Alternatively, decreased administration of the agent may  
be desirable to decrease expression or activity of  
Tango-77 to lower levels than detected, i.e., to decrease  
5 the effectiveness of the agent.

### C. Methods of Treatment

The present invention provides for both  
prophylactic and therapeutic methods of treating a  
subject at risk of (or susceptible to) developing or  
10 having a disorder associated with aberrant Tango-77  
expression or activity. Alternatively, disorders  
associated with aberrant IL-1 production can be treated  
with Tango-77. Such disorders include acute and chronic  
inflammation, asthma, some classes of arthritis,  
15 autoimmune diabetes, systemic lupus erythematosus and  
inflammatory bowel disease.

#### 1. Prophylactic Methods

In one aspect, the invention provides a method for  
preventing in a subject, a disease or condition  
20 associated with an aberrant Tango-77 expression or  
activity (or aberrant IL-1 expression or activity), by  
administering to the subject an agent which modulates  
Tango-77 expression or at least one Tango-77 activity.  
Subjects at risk for a disease which is caused or  
25 contributed to by aberrant Tango-77 expression or  
activity can be identified by, for example, any or a  
combination of diagnostic or prognostic assays as  
described herein. Administration of a prophylactic agent  
can occur prior to the manifestation of symptoms  
30 characteristic of the Tango-77 aberrancy, such that a  
disease or disorder is prevented or, alternatively,  
delayed in its progression. Depending on the type of  
Tango-77 aberrancy, for example, a Tango-77 agonist or  
Tango-77 antagonist agent can be used for treating the

- 92 -

subject. The appropriate agent can be determined based on screening assays described herein.

## 2. Therapeutic Methods

Another aspect of the invention pertains to  
5 methods of modulating Tango-77 expression or activity for  
therapeutic purposes. The modulatory method of the  
invention involves contacting a cell with an agent that  
modulates one or more of the activities of Tango-77  
protein activity associated with the cell. An agent that  
10 modulates Tango-77 protein activity can be an agent as  
described herein, such as a nucleic acid or a protein, a  
naturally-occurring cognate ligand of a Tango-77 protein,  
a peptide, a Tango-77 peptidomimetic, or other small  
molecule. In one embodiment, the agent stimulates one or  
15 more of the biological activities of Tango-77 protein.  
Examples of such stimulatory agents include active  
Tango-77 protein and a nucleic acid molecule encoding  
Tango-77 that has been introduced into the cell. In  
another embodiment, the agent inhibits one or more of the  
20 biological activities of Tango-77 protein. Examples of  
such inhibitory agents include antisense Tango-77 nucleic  
acid molecules and anti-Tango-77 antibodies. These  
modulatory methods can be performed *in vitro* (e.g., by  
culturing the cell with the agent) or, alternatively, *in*  
25 *vivo* (e.g., by administering the agent to a subject). As  
such, the present invention provides methods of treating  
an individual afflicted with a disease or disorder  
characterized by aberrant expression or activity of a  
Tango-77 protein or nucleic acid molecule. In one  
30 embodiment, the method involves administering an agent  
(e.g., an agent identified by a screening assay described  
herein), or combination of agents that modulates (e.g.,  
upregulates or downregulates) Tango-77 expression or  
activity. In another embodiment, the method involves

- 93 -

administering a Tango-77 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant Tango-77 expression or activity.

Stimulation of Tango-77 activity is desirable in situations in which Tango-77 is abnormally downregulated and/or in which increased Tango-77 activity is likely to have a beneficial effect. Conversely, inhibition of Tango-77 activity is desirable in situations in which Tango-77 is abnormally upregulated and/or in which decreased Tango-77 activity is likely to have a beneficial effect.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

#### EXAMPLES

##### Example 1: Isolation and Characterization of Human Tango-77 cDNAs

Cytokine genes IL-1 $\alpha$ , IL-1 $\beta$  and IL-1ra have been found to be closely clustered on chromosome 2, i.e., IL-1 $\alpha$ , IL-1 $\beta$  and IL-1ra are located within 450 kb of each other. BAC clones containing IL-1 $\alpha$  and IL-1 $\beta$  were used to identify other proximal unknown cytokine genes. To do this, a BAC clone containing IL-1 $\alpha$  and IL-1 $\beta$  was selected from a BAC library (Research Genetics, Huntsville, Alabama) using specific primers designed against IL-1 $\alpha$  and IL-1 $\beta$ . The DNA from the BAC was extracted and used to make a random-sheared genomic library. From this BAC library, 4000 clones were selected for sequencing. The resulting genomic sequences were then assembled into contigs and used to screen proprietary and public data bases. One genomic contig was found to contain two

- 94 -

segments of sequences which resemble IL-1ra. These two segments are potential exons of Tango-77 gene.

Two PCR primers were then designed from the two potential exons and used to screen a panel of cDNA  
5 libraries for the expression of a Tango-77 message. A cDNA library from TNF- $\alpha$  treated human lung epithelia showed a positive band of the predicted size (i.e., if the two exons are spliced together). Using the PCR fragment as a probe, a single cDNA clone was isolated  
10 from the same library. This cDNA contains an insert of 989 bp. The cDNA clone contains three possible open reading frames. The first open reading frame encompasses 534 nucleotides (nucleotides 356-889 of SEQ ID NO:1; SEQ ID NO:3) and encodes a 178 amino acid protein (SEQ ID  
15 NO:2). This protein may include a predicted signal sequence of about 63 amino acids (from amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID NO:4)) and a predicted mature protein of about 115 amino acids (from about amino acid 64 to amino acid 178 of SEQ ID NO:2 (SEQ  
20 ID NO:5)).

The second putative nucleotide open reading frame encompasses 498 nucleotides (nucleotides 389-889 of SEQ ID NO:1; SEQ ID NO:6) and encodes a 167 amino acid protein (SEQ ID NO:7). This protein includes a predicted  
25 signal sequence of about 52 amino acids (from amino acid 1 to about amino acid 52 of SEQ ID NO:7 (SEQ ID NO:8)) and a predicted mature protein of about 115 amino acids (from about amino acid 53 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:9)).

30 The third open reading frame (nucleotides 372-889 of SEQ ID NO:1; SEQ ID NO:10) encompasses 408 nucleotides and encodes a 136 amino acid protein (SEQ ID NO:11). This protein includes a predicted signal sequence of about 21 amino acids (from amino acid 1 to about amino  
35 acid 21 of SEQ ID NO:11 (SEQ ID NO:12)) and a predicted

- 95 -

mature protein of about 115 amino acids (from about amino acid 22 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:13)).

Tango-77 is predicted to be 35% identical to human IL-1ra at the amino acid level.

#### Example 2: Expression of Tango-77 mRNA in Human Tissues

The expression of Tango-77 was analyzed using Northern blot hybridization. A PCR generated 989 bp Tango-77 product was radioactively labeled with <sup>32</sup>P-dCTP using the Prime-It kit (Stratagene; La Jolla, CA) according to the instructions of the supplier. Filters containing human mRNA (MTNI and MTNII: Clontech; Palo Alto, CA) were probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Tango-77 mRNA was not detected in any unstimulated tissues (brain, liver, spleen, skeletal muscle, testis, pancreas, heart, kidney and peripheral blood leukocytes) mRNA on Clontech Northern blots.

Over 96 cDNA libraries were then tested for the presence of Tango-77 using PCR amplification. Only three libraries displayed a positive signal. These libraries were the TNF $\alpha$ -treated bronchoepithelium, TNF $\alpha$ -treated SSC cell line and anti-CD3-treated T cells.

#### Example 3: Characterization of Tango-77 Proteins

In this example, the predicted amino acid sequence of human Tango-77 protein was compared to the amino acid sequence of known protein IL-1ra. In addition, the molecular weight of the human Tango-77 proteins was predicted.

The human Tango-77 cDNA (Figure 1; SEQ ID NO:1) isolated as described above encodes a 178 amino acid protein (Figure 1; SEQ ID NO:2) or a 167 amino acid

- 96 -

protein (Figure 1; SEQ ID NO:7) or a 136 amino acid protein (Figure 1; SEQ ID NO:11). The signal peptide prediction program SIGNALP Optimized Tool (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that

5 Tango-77 includes a 63 amino acid signal peptide (amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID NO:4)) preceding the 115 mature protein; or preceding the 115 mature protein (about amino acid 52 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:8)); or preceding the 115

10 mature protein (about amino acid 21 to amino acid 136 of SEQ ID NO:11;SEQ ID NO:12).

As shown in Figure 2, Tango-77 has a region of homology to IL-1ra (SEQ ID NO:14).

Mature Tango-77 has a predicted MW of about 13 kDa

15 and the predicted MW for the immature Tango-77 is 19.6 kDa, 18.5 kDa or 15.2 kDa, not including post-translational modifications.

#### Example 4: Preparation of Tango-77 Proteins

Recombinant Tango-77 can be produced in a variety

20 of expression systems. For example, the mature Tango-77 peptide can be expressed as a recombinant glutathione-S-transferase (GST) fusion protein in *E. coli* and the fusion protein can be isolated and characterized. Specifically, as described above, Tango-77 can be fused

25 to GST and this fusion protein can be expressed in *E. coli* strain PEB199. Expression of the GST-Tango-77 fusion protein in PEB199 can be induced with IPTG. The recombinant fusion protein can be purified from crude bacterial lysates of the induced PEB199 strain by

30 affinity chromatography on glutathione beads.

- 97 -

Example 5: Alternatively spliced forms of IL-1ra and  
Tango-77

Computer program Procrustes (Gelfand et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:9061-9066) is an alignment  
5 algorithm that predicts the presence of alternatively  
spliced exons for a protein of interest in a stretch of  
genomic DNA. Using the IL-1ra sequence, Procrustes was  
used to search for the presence of additional sequences  
that might encode for alternatively spliced forms of IL-  
10 1ra in the two overlapping BAC genomic sequences (see  
Fig. 3 and Fig. 4). Potential sequences that encode  
variant exons for IL-1ra were identified. These  
predicted exons aligned well with the N-terminal region  
of IL-1ra, but were not present in Tango-77. The results  
15 from Procrustes predicts the existence of more spliced  
forms of IL-1ra.

Furthermore, Procrustes also predicted an  
additional sequence in BAC1 and BAC2 that encodes an  
alternatively spliced exon for Tango-77 (T77-procrustes;  
20 Fig. 5). This predicted splice variant form of Tango-77,  
T77-procrustes, was aligned with Tango-77 (Fig. 6) and  
with IL-1ra and IL-1 $\beta$  (Fig.7).

PCR primers within this sequence can be used to  
generate a product that can be used to screen a panel of  
25 cDNA libraries using standard techniques. Suitable cDNA  
libraries include libraries made from TNF $\alpha$ -treated  
bronchoepithelium, TNF $\alpha$ -treated SSC cell line and anti-  
CD3-treated T cells. The resulting cDNA clone(s) can be  
isolated from the library and sequenced to identify  
30 additional Tango-77 cDNAs.

- 98 -

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific  
5 embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

- 99 -

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

- a) a nucleic acid molecule comprising a  
5 nucleotide sequence which is at least 45% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof;
- 10 b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof;
- 15 c) nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the  
20 plasmid deposited with ATCC as Accession Number 98807;
- d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID  
25 NO:12, SEQ ID NO:13, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or the polypeptide encoded by the cDNA insert of the plasmid  
30 deposited with ATCC as Accession Number 98807; and
- e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9,

- 100 -

SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

10 a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10 or the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof; and

15 b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell containing the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

- 101 -

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

5 a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID  
10 NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, or SEQ ID NO:13.

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8,  
15 SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule  
20 comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the complement thereof under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is  
25 at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID  
30 NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807.

- 102 -

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

5 12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID  
10 NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807;

b) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID  
15 NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the fragment comprises at least 15 contiguous amino acids  
20 of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807; and

25 c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the  
30 plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid sequence of

- 103 -

SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10  
under stringent conditions;

comprising culturing the host cell of claim 5  
under conditions in which the nucleic acid molecule is  
5 expressed.

13. A method for detecting the presence of a  
polypeptide of claim 8 in a sample, comprising:

- a) contacting the sample with a compound which  
selectively binds to a polypeptide of claim 8; and
- 10       b) determining whether the compound binds to the  
polypeptide in the sample.

14. The method of claim 13, wherein the compound  
which binds to the polypeptide is an antibody.

15       15. A kit comprising a compound which selectively  
binds to a polypeptide of claim 8 and instructions for  
use.

16. A method for detecting the presence of a  
nucleic acid molecule of claim 1 in a sample, comprising  
the steps of:

- 20       a) contacting the sample with a nucleic acid  
probe or primer which selectively hybridizes to the  
nucleic acid molecule; and
- b) determining whether the nucleic acid probe or  
primer binds to a nucleic acid molecule in the sample.

25       17. The method of claim 16, wherein the sample  
comprises mRNA molecules and is contacted with a nucleic  
acid probe.

- 104 -

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which  
5 binds to a polypeptide of claim 8 comprising the steps of:

- a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and
- 10 b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- 15 a) detection of binding by direct detecting of test compound/polypeptide binding;
  - b) detection of binding using a competition binding assay; and
  - c) detection of binding using an assay for
- 20 Tango-77-mediated signal transduction.s

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a  
25 sufficient concentration to modulate the activity of the polypeptide.

- 105 -

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

- a) contacting a polypeptide of claim 8 with a  
5 test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```

GTCGACCCACGCGTCCGCAGACGTCTACCTGGGGGTCCCGTCTGCGCTCCCGGGATGGAACGCCCAGGGGAACTTA 79
GGCAGGCGAGCGGACGGGCACCTCCCGCGGGACGAACTCACTCGGTGGCCTCCTACTTCCCCGGCCGTGTTCCAACGCC 158
TGAGAATAACGGGAACAGCGGTCTACTCACCAGACAGCGGCAGCAGCGGCCTCTCTCAATTGGGCAAAGCACTCCAGAC 237
CTTTTGGGAAGAGTGACACCAAAGGCAAGCACCTGCTTGGCAGGCCCCCTCAGCTTCTACGCAAGTATAAGTCTTGGACTT 316
CATTCATTTTCTGTTGAGTAATAAACTCAACGTTGAAA M S F V G E N S G V 10
ATG TCC TTT GTG GGG GAG AAC TCA GGA GTG 385
K M G S E D W E K D E P Q C C L E D P A 30
AAA ATG GGC TCT GAG GAC TGG GAA AAA GAT GAA CCC CAG TGC TGC TTA GAA GAC CCG GCT 445
G S P L E P G P S L P T M N F V H T K I 50
GGA AGC CCC CTG GAA CCA GGC CCA AGC CTC CCC ACC ATG AAT TTT GTT CAC ACA AAG ATC 505
F F A L A S S L S S A S A E K G S P I L 70
TTC TTT GCA TTA GCC TCA TCC TTG AGC TCA GCC TCT GCG GAG AAA GGA AGT CCG ATT CTC 565
L G V S K G E F C L Y C D K D K G Q S H 90
CTG GGG GTC TCT AAA GGG GAG TTT TGT CTC TAC TGT GAC AAG GAT AAA GGA CAA AGT CAT 625
P S L Q L K K E K L M K L A A Q K E S A 110
CCA TCC CTT CAG CTG AAG AAG GAG AAA CTG ATG AAG CTG GCT GCC CAA AAG GAA TCA GCA 685
R R P F I F Y R A Q V G S W N M L E S A 130
CGC CGG CCC TTC ATC TTT TAT AGG GCT CAG GTG GGC TCC TGG AAC ATG CTG GAG TCG GCG 745
A H P G W F I C T S C N C N E P V G V T 150
GCT CAC CCC GGA TGG TTC ATC TGC ACC TCC TGC AAT TGT AAT GAG CCT GTT GGG GTG ACA 805
D K F E N R K H I E F S F Q P V C K A E 170
GAT AAA TTT GAG AAC AGG AAA CAC ATT GAA TTT TCA TTT CAA CCA GTT TGC AAA GCT GAA 865
M S P S E V S D * 179
ATG AGC CCC AGT GAG GTC AGC GAT TAG 892
GAAACTGCCCCATTGAACGCCTTCCTCGCTAATTTGAACTAATTGTATAAAACACCAAACCTGCTCACTAAAAAAA 971
AAAAAAAGGGCGCCGC 989

```

Fig. 1

```

1      50
MEICRGLRSH LITLLFLFH SETICRPSGR KSSKMQAFRI WDVNQKTFYL
~~~~~
T77-human ~~~~~
IL1b-human ~~~~APVRSL NCTLRDSQQK SLVMSGPYEL
Consensus -----

51      100
RNNQLVAGYL QGPNVNLEEK IDVVPIEPH. ALFLGIHGK MCLSCVKSGD
~~~~~
T77-human ~~~~MNFVHT KIFFALASSL SSASAEKGS. PILLGVSKGE FClyCDKDKG
IL1b-human KALHLQGQDM EQQVVFMSF VQGEESNDKI PVALGLKEKN LYLSCVLKDD
Consensus -----LG-----L-C-----

101      150
ETR..LQLEA VNITDLSENR KQDKR.FAFI RSDSGPTTSF ESAACPGWFL
QSHPSLQLKK EKLMKLAQK ESARRPFIFY RAQVGSWNML ESAAHPEWFI
K..PTLQLES VDPKNYP..K KKMEKRFVFN KIEINNKLEF ESAQFPNWI
Consensus -----LQL-----F-F-----ESA--P-W--

151      192
CTAMEADQPV SLTNMPDEGV MVTKFYFQED E~~~~~
CTSCNCNEPV GVTDKFENRK HI.EFSFPV CKAEMSPSEV SD
STSQAENMPV FLGGT.KGGQ DITDFTMQFV SS~~~~~
Consensus -T-----PV-----F--Q-----

```

FIG. 2

## &gt;Contig1

GAAGTGAAGATATAATGTATAGTAGTAATATATAATGTTAGGTGAATTAA  
AGGAAATAGAAATATATTGGGGAGTAATTATGGGTGTAAAGAAATATAGTA  
GGGAAGTATTTAGATTTGAGAAAAAAAAGGAATTTAGTGTAGGTGAA  
NAATAAAAGNANAAGGTTAAAAATTAAAAAAATTTAAATATAAATAAAT  
AAATAAAATAAAAATAAAATAAAAAATTTAAAAAATTTAAAAAATATAA  
AAAAATAAGAAATGGAAGTGGATTCTTAGAAAAAAGAAAGTAAGGTGA  
TATGAGGAGATAGAGAGGATGTGGTGTGAGATGATTGGTTTAATTAGAAA  
ATAGGTTTTGAATAGAGTGGGAAAGTAGAGTTTGGTAAATGTGGGGGA  
AGAGGGTAATGTTGTTTGAAGTGAAGAAAAAATGGTATATTTTTATAAAA  
TAATGAGGAAAGTGTGTGAAAAAAAATTTATGGGATTTGGGAAGGTGAT  
ATATAAAGTTGTGGAAAAATTTGGGGGGTGGGGTTTATTTAGGATTAAAAA  
GTTATTTAAGAATGAAATGAATTTTTGTTTGTAAATTTGGGGATAAGAA  
ATTAATGTTTAGAAAGAAAGGGAAAAAATTGAAGAAAAAATTTAGATTT  
TGGAAATTTAAAAATATTGTGGGTGTAAATAGGAAGGATTTTAAAGGTA  
ATTGTGGAAGGGATTTGTGTGGAATAATAGGGAGAAAAAATGGGG

## &gt;Contig2

GCATCTAACTGGAGCCTGCATTATTACAGATTTAGCATCACCAAAGTCTA  
AACAATTAGACTGACTAAGGCAGAACTGCCCTTATGACAGCAGACATAAG  
AAGGAAAAGGCCAAAACACTGTGTTAAAAATTATCCAAATGTGAGGAAAA  
GGCAAAGAGAGTAGGTGTGCCTTTTTAGTGTCTAAGCTGCCTGCCCAAGG  
GGCATCTGATGCTCTCAGGCAGGAGTCCACAAATTTTTTTTTGTAAAAGA  
TCAGATAGTAAATCTTTTCAGCGTGAAGAGCATGAGGTCTCTGTCAAAA  
TACTCAACCACCATACAACATGAAAGCAGCCAACAGACAACACATGACA  
AATGAGTGTGGCTGTGTTCCAGTAAATCTTGATTACAAAAACAGGCAAGA  
GGCCAGAGCTGACCCATGGGCCATAGTTTGCTGACCCCTTCTGTAAAGGA  
AAGTATTTTTGTTTGAAGTGTGTTTACCATTGATTGAACACAAGGCTCT  
GTAAAGTTACTTGTAACTTGCAGAAGATTGATGAGTGGCAAGTAATTTT  
TATTCACCAGAATATAAAATTATTTCTGTTCAAGTAGAAAAGATAAACCAA  
CTGTGATATTATGGTCCTG

## &gt;Contig3

GGGGTGTCTGTCTACCATGTGCTCGCAGTTCTGTAATAAATGTTCTCTCA  
AGATCCTTAAATCTCTTGAAATTATAAAAAATATTGGAAAGAGAAGAAC  
AGTTTTTAAATATATATATATATATATTTTTTTTGGAGATGGAGTCTT  
GCTCTGTCTGCCAGGCTGGAGTGCAGTGGCGCAAACCTTGGTTCACCACAA  
CCTCTGCCTCCCGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTGAGTAG  
CTGGGACTACAGGCGCCCGCCACCACGCCAGCTAATTTTTGTATTTTA  
GTAGAGACGAGGTTTTACTATGTTGGCTAGGCTGGTCTCAAACCTCCTGAC  
CTTGTGATCTGCCCCGCTTGGCCTCCCAAAGTGTGGGATTACAGGTGTG  
AGCCACTGCACCTGGCCAGTTTTTTAAATATATTTTTTAAAAACACTTGAA  
TAAGAGTCAGTGTAACCTAGAAGTTTAAAAATGCTTCACAGAACACCCAG  
GGTTTACATTACAAGATTCTCACAACAAACCTATTGTAAAGGTGAGTAAG  
GCATGTTTATTACAGAGAAAAGTTTGGGAGCAAACTGTAAAAAATTATAT  
TTTTGTTGATTTTTCTAAGAGAAAGAGTATTGTTATGTTCTCTAACCTC  
TGTTGATTACTACTTTAAGTGATTTTCTTGAGAGCACATGATGATCC

## &gt;Contig4

GCCGTTTCATAGAAAACCTGAAAGCAATAAGATGACTAGGTAAGCATGACAT  
TAAAAGGTATTTCATGGGACGTGGTTACAAAACCAACTCACAACATAAAA  
GTCTTAGGACCTCTCGCTGACTTAGGAGCCTGATCCCAACTCTGAGAATG  
ACTCAGTGTGTTACCCTGTGGCTAGTGTAGACCAATGATCCTGTCTCAGA  
GTCACATAGCCAACAGCCCATATCAAGTACTTGAACTTTGACTCAGAAAC  
CTCAGTGTGAGAACCTTTGACCTAGGAACACCTGTAGTGGTTAACTGCA  
ATTTGCACCCCTTAGTTTCAGGGCTTTACAACACCGGGGGCGGGAGGGGA  
AAGGCATANANCTGATGACCTAAAGGAAACCCATTGCAGCAACCGTTTTG  
TGTTAAGTGTACAAATAAGTGTGTTTTAGAAATCCTCCAGGTAATGCCTT  
TGTTATTTAATGTGTCTGAGACAATTCTGCACATTAAAGAATATAAATA  
TTACCTTGTAAATCCAATTTGAAATGTGTAATTGACATTAGACTTCTATT  
TGAATTTGAAATGTCTAAACAATGTGGTTAAGTTTGTAAAGGTGTGTG  
AATTTTGAGTCTGATTTACTACATTTTTTTTTTAATTTCTTTTTTTTGG  
AGTTTTAGGGATTGCTTAGATGGCTAGAAAGATTTTATTCATCAGATTTT

TAAGTCTGCCTTGGCAGGCACCTTGCACTTTTGAAGAATCAGATATATC  
AAATTTGTAGTTTAAATATTTAAGGGAACCAATTAACATGCTAGAAA  
AGAGAATTAAGTATTTAGGAGGATTTAATATGGTGTGAAAGTTGTGAAAA  
TCAAAATGGAGACACTAATGTTAAGAAAACCCTGATAAATGGAACCAGGG  
AAAGGCATGAAGATAGAGTTCTCACACTTGTATCCCTGATCATGAAAAAG  
ATCTGC

>Contig5

GGGTTTTTCCGCGTTTTTACCCGAAATCTTCAAGGGATGGGAAAAAGAAA  
ATTGCTAAAAAATCTCGGTTTTTGGTTTTAACAGATATTTACACCNTGG  
ATCCCATTTATTATGTTGTCCCAAGTTTTTCGGTGGGTCCCAATCAGT  
TAGCCCCCTCCACAGTGAAAGCACTTTACTTTATCACCTTCACCTAAAG  
CATAAAATCCAGCTCTTGAAAGCTGCTCCTTGTTAACTGAATATATCCAC  
ATCCCAAAGTAATGATCCATGCTTCATAATCTGCCACGGATGGATGGAT  
GGATGGATGGATGGATGGATGGATGGATGAATGGATGGATTGATTTCTTG  
GAGGATTTGTTGAATTTGGGAAATTCACGCCAGGACAGCTGGCCCAAAC  
TGCCCCGCGACAATCTGCTCGGTACAAGGGGAGGGTCTGGAGAGGGTGCG  
GCCCCGAGCCCCAGTTTTGGAAATGCCAAGTTGGCTCTGCAGCCGGGCCTTA  
GCCACTTGGGTCTGGCGTCCCTCCATTATTAGCGCCATGCCGGCTCGGGG  
TGCTGCCAAGTCCCTGAGAGCAAGCC

>Contig6

CGCGCTCAAGAAAAGCTGAAGTGTGAATGTTCTGTCTACCTTCACAGTAA  
ATGCTAAGAGAATGACCCAAGAGCAGAGGGTATCACTCTGCTACGGAGGA  
TTGATTGTAAGTGGCTCTCCTGCCTTAGCAAGAAATGCCAGAACCATGGT  
CATTCAAGTTCTTGACCAAAACTGCCTTCATGAGAATCAACTTCCCCAA  
GAAAAAAAAGCAGAAACAGGCAAAGCTTCCAGCATGGTAGGTAATACTG  
ACCCTTCTTCCCTCCTTCTTGGAGATTACACAGTAATAATGCATAAA  
GCTTTGCCAATGGACTAAGCACTGCCCAGGGTTTTTGTATGCCTGGAC  
TGAAATGCTCTTTTTGCGTTATCATAGAATCCAGTGCAGTCTGAGTAGA  
CTCTAAGCAAAAGGGACATTTTTCAAAAAGGCTTTAAATTGCTAGTACAA  
AGAAGGCAACAAAACCTGCGTAACTGTGGACAGATTAACTCACTTGGTGT  
TTTGGCTCTTCAGTTTTCCCTTGGCTGCGAAGTACTCCTGAAGCTTTCTC  
TGCGGCTCTTCTGCAAGCAGGCAAGCAAAAAACGACTGAACCTTTATTT  
CGAGAT

>Contig7

GAAGAGCCGCTAACTTGCTGTAGTGATAAGGAATGAACTAAGGCTAGGGA  
CATATTAACATCCGCTGGTGGTGAATCTTTAGCCTAGATCTTACCCCACT  
CCTGCTCCTTCCATATGGTTCGGTCTCAGGCTCACTACCGATCAATGGCG  
TACTAAAAGCACTAATATAGACTCCAACACGTCTGTCTGTGTTTCAGC  
ACAAGCCGTGGAGTTAATCCCTCTGACAGTAGCTCAGATAAGGATGGGCT  
ATCATGGGCCCCGGAAGTGGGGCATGACGCTCGTCACCAACGCATGAGCTC  
CCCAAGTATGCTATACCTGTCCCTATGAAGGGCTTCCAACCTATGTGCA  
GTCCCCATGTGGAGAGTCAGGTATTGATTGATCAAGCCAGGGGTGTGGTG  
AATGGGGAGCTTCTTACAGGGGTAATGATAATTGAAATGCACGGTGATGG  
GGATTTTCATATTGGTCTCCTAAGGAGATAACAGATTGGATGCGGGGTGCG  
ATATTCCACTGCCCAGGGTGTGTACCGAGGGTATCTGCAGGTGGATCTCC  
TCCCCACGTTTGATTAATACTCCTGTCTTGGGAAGCATAGACGGGCGGGG  
GAAATGATGAAGGGTGACCACTCCCC

>Contig8

GGGAACGCAGTGCTCTGTACGATGGCCTTGATTGCGAATTCCTGCAGGGG  
GGG

>Contig9

GGCAAGAGATTTAATATTCATTCCATCTTCATTTGGAAGATGAAAAATTG  
GGGACCAGAGAGGGGAGGGGACTGGGCCAAGTTTTCAAAGAAAAGTCAGT  
AGGAATTGTGAATTCCTGGGGGCCGGGGCCATTAGTGCTGTTTTGGATC  
AGTAAATGGAGATGTGAGTTTCAACAGTAACAGGGACATTTTAAATTA  
AATGATTTAACCTTTAGAAAATGTCCTATTTTGTAAATAATGATGGATTCA  
CAGGAAGGTACAAAGAAATGTCCAGAGAGTTCNTGAGCCCCCTTCAGCCA  
GCTTCTTCCAATGTTAACATCTTGCATTATTATAGTACAACATCAAACT  
GGGAAATCGATATTGGTACTGTCCAGATAGCTTACTCAGATTTTGCCAGT  
TATACTTCCACTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT

TGTGTGTAGCTCTATGCAATTTTATG1...GTAGCTTCATGTAAACACCC...  
AATCACAATACTTAACCTATGCCCTCATCACAAGACTCTCTCTTGCTATGC  
TTTACAGCTGTATCCTCTTCATCTCCAAACCTAAGCCCACCTCACCGCC  
TCCACCATCTCTAATCCCTGGCAACCACTATTCTGTGCTCCATCTCTGTA  
ATTAATTGTGTTAATTAATGTTATACAAATGGAATCATGAAGTATGTGTC  
CTTTGAGATTGGGCTGTTAATTTTTCTACTCAGCACAATTTCCGTGAGTCT  
AATCCAACCTTGTGTGTAGCAGTAATTTCTTCTTATTATTGCTGAATAAT  
ATGCCATGGTATGGATGTATCACAGTGTGTCTAATCCTTTGCCCATTGAA  
AGGAATTTGGATAATTTCCAGGTTTTGGCTATTATGAATAAAGTGAACAT  
AAGACATGTGTGTACAAATTTTGGTGTGATCAAAAGTCTCATTTCTCTGG  
GATAAATGCCCGGTAATGAAATGGCTGGGTTGTGTGGG

>Contig10

GCAAGAACACAGGCGCTATTATAACCTTACTACCAAGACCTGAACCCAT  
ATAAAGGTTTATGCGTAACAATCATCATCCCTGTTCCAGAAGATTACACG  
TACGACCACGCCTGGCTCACCAGCTCACGTGGGCCAGTACCAGAAATTCT  
CCCAAACAACAGTCGTGTCTGAAAACAATCGCGGTGACCTCCACGGTTA  
GAAAAGCCTGTTTTCAAGTCTGGAATTGCCACATATTAGCTGGGTAACT  
TTGGGCATCACATTTACTCTCTCCGAATTCAGATTGCAAAAACCTCATTG  
GATTGTTTTGTGGATTGAAAGAAATAATGTAAATTTAGGCCGAGTGCTTT  
GACTTACGCCTGTAATCCTATCACTTTGGGAGGCCAAAGCAGGAGGGTCA  
CTTGAGCTCAGGAATTTGAGACCACCTCTGGCAACATAGTGAGATCCTGT  
CTCTACAAAAAATTTTTTTTAAATTATCCAGCATGGTGGTACACGCCTGT  
ATCCCAGCTACTCAGGAGACTGAGGTGTGAGGATTGCTAGAACCTGGGA  
GATCAAGTCAACAGTGAGCCGTGGTTGTGCCACTGCCCTCCAACCTCAGT  
GACAGAGGAAGACCTGTCTCAAAAAAAAAAAAAAAAAAGTAGTAAGTTTAA  
AGAACTTAGTGTAGGCCTGGCATATAAATGATATTGTTGATGTTGATGTT  
AGCTTGAAGGCACATTTATAGGAGTAGGGATTTTATAACATTATGAGCCT  
GAGAGCACATATAATGTTCCC

>Contig11

GGTCTAACATGCTCCAACCTGAAGAAACCCACACTTGTCCGGCAAGGAAA  
CTACTACAGATTTCTGACCTACTGTGCAATTCGGGGCATGCGACGGGAC  
TGTGTTTTCTGGGTACGCTGTCTCAGGTTCTGTCTGGGATGTAAGAATTCAA  
CTTCAGTAGTTCTCTCATAGACGCCGACGAGAGGGGGCGTCTCTTTTCTCT  
GATGAATCTGCCAGATCTTCCACTTCATAGAGTCTAAATCCTCCGATTCTG  
ATCTACTGGAGACCCCCACGTTACAAAACGTTAAACGTCGGTGACAGCT  
CCCCACATAGGGAAAGATCACCTGAGTCTCACTACCTCACATTAGTGCTA  
TCTCCAGCCCCATGCTATCTACGAGATGGTCACGCGAGGTTTAAGGGGTC  
TCCGATTCCGGTGGTCCGATTCTAGCTAATCGTGGCCCTACGTGAACGATC  
ACTCCTGCTCGTAACATCGATACAGGGTTCGCGCTGACAAATGGTACTACG  
TAGGTTCTCAGGTCAATGCCCGCTCACGAATGAGCCTAACTACCCCATAA  
GTGCACGTACTGTGTTACCTTCTCTGTTCCGGCCAAACCTGCTACTGTATG  
CTGTGCTTGT

>Contig12

AGGCTCCATGTGCTCTAGCCTGATTATCTTTTCAAGTGTTTTATTTGCTA  
ATCTATAAGGCCCTTTTCGTAAAATGTTCACTCATTTTCTAATTAGATAT  
TTTTTTTAAATGTTGAGTTTTGAGAGTTCTTTAGATATTTTAGATACAAGT  
CCATTGTCAAATATGTGATTACAAATATTTCTCTCAATCTGTAATTTA  
GTTTTTCATCTCTTAACAGGGTCTTTTGGAGAGCAAATAATTTGATTTTC  
ATAAGGTTCAAATTATTAATTTTTCTTGTATAGTTCACTTCTAGTGT  
TAAGTCTAAAACTGTGCCTTGTCTAGGTACCAAAGTTTTCTCCAGTT  
TTTTTTCTAGAAGTTTAGAGTTTCATGTTTTACATTGGAGTCCATGATCC  
ATTGTTAATTAATTTTTGTATATAGGTAGATGTTTAGGTTTAGGGTTTTT  
TTAAAAAAAATACATATGTTTAATTGCTCCAGTTCCTTTTCATTGAAA  
AGGGTACTCTTCTCCATTGAATTGCCTTTGTGAGAAATTAATTGGACAT  
ATTTGTGTGAGTCTATTTCTGGGCTCTTTATCATGTTACTTTTAAAAAAT  
GCATCAGTTCCTCCACCAATACCTCATTGTCTTGATTATTGCAGTTATAT  
AGTAAGCCTTAGCATTAGGAAAAGTGTTTTTCTGCTTTATTCTTTNTCA  
AAAAATTTTGGATATTCTAGGGCCTTTACATATAAATTTTAAAAATACT  
TTGTCTATGTCTAACCGAAAGCCTTATGAAGATTTTGATAAGAATTGCAT  
TATGCCTATACATTAATTTAAAAAGAACTGATGTCTTTATTAGTTGATT

FIG. 3 (3 of 52)

5/118

CTGCTAATCTATGAACAAGCATCTCTCAAAGCATTAGTCTTTCTT  
AATTTCTGTCAATATTTTTTAAATTTTCATCCTAAAGATTCTGTATAT  
GTTTTGTTGAATTTATGCTTAAGCATTTCACCTTTCTTGGTAACAATTATA  
AATGATTTTGTGTTTTTATTCCACTAGTTCATTTTCAGTGTGTAGAAAA  
GCAATGAATTTTGTGTGTTGATCTTTGTTCCCTACATCTTGCAACATTAT  
TGAACCTCATTATTAGTTCTAGGAGGTTTTTTCATTTTCTTGTAGATAC  
CTTGAGATTTTCTATATAGACAGTCATGTTGTCTGCAACAGGCACAGTT  
TTATTTCTTCCTTTTCAATCTATATGCCTTTTTTTTTTTTTTGCCTTAT  
TGCAGTGGCTAGAACTTCTAGCACTATGTCAAATAGCATTGGTGAAAGCA  
GACATCCTTGTTCTTGTCTTAGAGGAACATTTGGTCTTTAATCTTGGAT  
TGCG

>Contig13

GCGCCTCCTTTCTCTTCCAAAATTTCTCTTGTCTAGTTATTTGTCCAGG  
GAAATTTGAAAGCTCACTTACTGTGCAAGTCAGCAGGAAACAACCTGGGTC  
TGTGCACAGCACCTAGCAAAGTTCTGCTCTAGGAATTACACTTTGGCCCT  
GAGGTAGATTTCTACAAGAACCTTACCTTCTAAGCAGCACTGGGGTTCAT  
CTTTTCCCAGTCCCTCAGAGCCCATTTTCACTCCTGAGTTCTCCCCACA  
AAGGCATTTTCAACGTTGAGTTTATTACTCAACAGAAAATGGAATGAAG  
TCCAAGACCTAAGGAGATAGAAAGGGGACCAGTTATGGCATCTTCTCACC  
CCAGGACACCTTGCTGCATGTCTCTAGTGTGAACAGACCACTGGCCTTG  
CTCTGTAGTTTGAATGCTCGCTGCAACCAGAAAGGCACCAAGGGGCCAG  
ACCATGCTCTCCTGTCTATCACGCCTTCAAAGCAGAATTTCCCAAACCTT  
GAGTCACAGTGCTAACACACGGGGTGCCATAACATTTTGTGATTGTTGG  
CATTTTACAAAATAAAAATAAAAAGTTAAAATGCATTGCTCTATTCTT  
GGGGCTGGCACACTATTGCCTTTGGCCAAATCCGGTCCCTGACTGTTTTT  
TTAAATAAAGTTTTATTGAAACACAACCATGCTCTTGTGTACATATTGTC  
TCTTGGCTGCTTCGAAGCTACAATA

>Contig14

GTGTTGCTTTTTTAACTTACCTAAAATTACTCTGTAATCCATGGATCC  
TTAATTTATTTAAAAAACTAATGTTAATGAGTAGCTTTATTTTCTCCCA  
TCTAATTTAAGGGCCACAGAACACCTTCACTTACCTCAATCCTCTCCCAA  
CTTACATGCTTTTAAATGTCATATATGTTAATACCGTATACTTTTAAACT  
TTCTAAAATAGCATTATTTTATAGCATGAGTGTTCAATTTACATTTTGTGA  
TATATTTAGAATTTTCTTTGCTCTTCGTTTCTTCTTCTATTTATGACTCC  
CCTCTGGGATCATTTTCTTCTACTTGAAGTACATAGTTTAGAACTGCAC  
TATTCAATACAGTAGCCACTAGCCATGTGTAGCTATTGAAGTTTAACTA  
AGTAAAATTGAGTAATATTAAAAACTCAGTTCCTTCATCTCACTAGCCAC  
ATTTCAAGTGCTCAGCAGCCACGTGCGACTAATGACTACTGTACATCAA  
CATATAGAACATTTCCATCATGGCAAAGAGCTCTATTGATAGTGTTTCATC  
CAGAGTTTCTGTTCCAGGACCAAACCTGAGGGTTGGGCTGCTATTTCTCAT  
GGCCCAATAACAAGATGCAGATGAGCTGGGGAGGAAGAGAGTTTTTATTT  
CTGCNACCATTTACCGGGAGAAGGCCTGGAAATCATCACCAGGCCAACTC  
AAAATTATTACGTTTTCCAGAGCTTATATACCTTCTAAGCTATATGTCTA  
CGTGTAAGTGTGCATTCACCTGAAGACGTTAGTGATTAACTTCTTTTAAAT  
CTGTAACCTAAGGTCTGAGTCCGGAAGATCTTCCCCTGGAGCCTCAGTAAA  
TTTACTTAATCTAAATGGGTCCAGGTGCTGGGGTAATTACCCTTATCTTG  
TCCCCTGCTAAATCATGGAGGTTTGGGGATTCTTTAGAGCACCAATAAA  
CTTGTTTGTGGAGGCCTGGGGGTTTCTTCTGACCCACAATAAACTTGTT  
TAATCCTAAATGGGTCTGTTAAGAATTCCTTCTTTATTTTGTATATTT  
TAAGGCCCAGAAAAGGCCTGGGCAAACTCTTGATGGGCTTTTGTACAT  
TCCAGCCTTTGTATAAGAACTGGTTTTTAATATTTAACTTAACCATTT  
AGTCAGTACTGAAACAGTTGTTATAGAGATCTGCATTAGTGAGACCTGGC  
CTGCCACATTTCTTTTCTGAAGATCTTATGGTAGTGATCACCTTTGTGA  
AAGGAAAAATAATCTTGGGACCTCAAAATCACTAAGCCAAAGAAAAAAGT  
CAAGCTGGGAAGAACTGACACTTAAATCCAACACTGCTAACTCATTTCAT  
CTCACTCATTTCATTCTTTTATTTTCTTTTTTCTTTTCTTTTTTTTTTTT  
TTTTTTGAAACGAAGTCTTGCTCTGTCAACCAAGCTGGAGTGCAGTGGAT  
CTCAGGTCACTGCAACCTCCACCTCCCGGGTTCAAGCGATTCTCCTACCT  
CAGACTCCTGAGTAGCTGGAATTACAGGCACCTGCCACCACGCCTGGCTA  
ATTTTTATATTTTATAGTAGAGACGGGGTTTACCATGTTTCATCAGGCTGG

TCTCGAACTCCTGACCTCGTGATCCGC...CCCCCTCGGCCTTGTTTGCT.  
GAGGTACTGTCTAAATGCTGGAAGTGAAGTGGCAAGCAAGACATCCCTA  
CCCTTGAGGAACTGTAATCTAGTCGGAAATACAGATGTCAACCAAGTCT  
CACACAAGAAATTTGTACAAAACCCCTAGGA

>Contig15

GGAAAAACCTATCACCGCCTCCTATGGAACCTTAAACAAAAAGAAAAGTA  
ACAAAGGAAATGAATATTTTCATTCTGGAAGAACATTGAAAAAGAACAGGA  
AGAAAGAGAAAGCACAACTCGAACTGTCCACTAGAAATTGACAACACTCTGA  
CAGAATGTCTGAACCTCATCGAAGGGGTAAGTGAAAAAATAAGCTCCTC  
CAGCTTTGGCCCCAAAGTCTTATAATTTTAAACATATTCTTAAATATAAT  
ATAGGAGAGATAGCCTTCATCTAAGTAGAAATTTAGCTACTCTTGTAAT  
ACAGAGTAATAATAATGACATGCCCATAAACAGTGTCTTTTGTGTAT  
CTGTGCTTTTATAAGCACTTAGCTAAGATTATCTCACATAATTATCATAA  
CCACTGTTACTATGACCACTTTACAAACAAAACCTGAGGCACAAAGAAGTT  
GGAAAACCTAATCCAAACAACTGGCTCCAAAAGGAACCTTGCTTTCTTTG  
GGTATCAAGTTCTGAAGAGTACACATTTAACATTGAACTGAGGTCAGAA  
GGCAAGTTTCTATGTAAAGTTGGAGTATTCTGAATACTCTGGGTAGCTAC  
AAATAGTATTTAAATTTTATCTTTGGATTCTGCAGATAAGGATAAAATAGA  
TGGTAGGCAAAGAGTATGATCCTTAGGAGAAATTTTCTGAAGGAAAAA  
TATATTAATAAAAAATGATGGAATAAACTTCTAAGATCCTTGCCCTAGAGC  
AAAACCTCATTCACTCCTTTGGCTGGTAATGTTGAACATCAACAAAAA  
GGAAAAGTTTCACTTTAAGTCTACTCCAGGCAACATTTTCACAACATCCAG  
TTAAATATTAACCTATTTCTCTTTGTGGAATTGAACTAGAGTTCTTTTCT  
TATCCTCTTTTGGTTGTTGTATTATTTAAAAATGAGTACCTTTTATT  
ATTGAAATCATTCAAGTAATGCAGATAAATGATCAGCCCTCTCCCTGTA  
CAAACATACATACTTAGGCATCCCAAACCTTCTCTCTGGAGGTGACCACCA  
TTGCCAGTCATTCTGTTTTCATGCATGTCCATACAGTATAGGTATG  
TCGAGAAATGAAGTATTATTTTGTGAGTTGCAATTCTTTTATTACACA  
TTTTTGTGTAATTTGGTTGTCTTTCTTGTGTTTCTTAGTACCAATGTT  
ATGCTGACTTAGGCAGATGAGTTGAGTATTTTCTTTTGGCCCTATAAAC  
TGAAAAATAGTTTGTATGACATGAGAATTATTTTATTTTGAAGGTTG  
ATAAAAACTTGCCCATAAAAATCGTCTGGACCGTTTCTTGAGGATGCCT  
GTGTTAGAGCC

>Contig16

CGCTTTAACCTGGGCTACCAATGGTTCGTCAAGTTCTAGATTCTCTATTA  
ATACCTTTTTCTTGTGTCTTTCTCTGGTCTGTTTTAGCCCCGAGTCTCT  
TAGATCTGTCTCTAATATTCCTATTGACTTTACTTCATTTCTAAGTCT  
TTATCCTTTTGCTTTACTTTCCGAGAGACCTGCTTAACCTTATCTCCAA  
CTCTTTTATTGAATTTCACTTTCTTTTACTATATATTTTACTTTGAATA  
CACCTCTCTCTCTCCTCACATTTTCCCCCATAGTATTTTGTCTTCAATTGA  
CAGTTCTACTACTTTATTACTCTGGAGATATTAATAAGTTTTTAAATT  
TTTATTTATTTTATTTTCAAAACAGTGTCTTACTCTGTCACTCAGCTG  
GAGTGCAGTGGTGTGATCATGGATCACTGCAGCCTTGATCTCTGAGCTCA  
AGCTATCCTCCTGCTTCAGCCTCCCAAGTAGCTGGAACCACAGGCATGTG  
TCACCATACCCAGCTAATTTTTTTGTTTTGAGGTGGAGTCTCACTCTGT  
AGCCCGGTCTGGAGTGCAGTGGTGCAATCTGGGCTCACAGCAACCTCTGC  
CTCCTGGGTCTGGTTCAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGC  
TGGGATTACAGAAACACACTACCATGCCAGCTAATTTTTGTATTTTGT  
AGAGACAGGGTTTACCATGTTGGCCAGGCTGGTCTTGAACCTCCTGACCT  
TGTGATCTGCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAG  
CCACTGCACCCGGCCACTAATTTTTAAATTGTTAATAAAGACGAGGTCTT  
GCTATGTTGCCAGTATGGTCTTGAACCTCCTGGGCTTAAGTAATCCTCCT  
GCCTCAGCCTCCCAAAGTGTGGGATTACAGGTGTGAGCCACTGAATCTG  
ACATTTTTTAAAGTTTTCTTCTCTTTACCAAGTCTTTTTTCCCCTTCT  
GCTTTTTTGGGTTGTTTTATTTTGATCTCTATCTTGCTAGAACTTTCTG  
CAGACGTTTAGTAATACTAGATTTTTGAGAGTGGGCAACTGGAAGCTGA  
TTGGAACCTCTGAATACATGGGTGAGGCTTGTGGCTGTGAGTGTCTTG  
CTTGATGTCCTGGCAAGGCCAATGGGTTTGGGACCCCTACTATTAGTATA  
GGCCTGATTTCCCTGGGAAAGGCTCTTTTGATCTCCTGCCTGGAGGATAAA  
GGCCTGGCTACCAGCCTTCTGTGTGTAATGTGAGGGAGAAGGGCTGGAGT

ATTCAACATCATGCTGAA.CCTTTCAA.JATCATCTTGTTTTTAGTAATC  
TCCTACCTTAACTCTCTGTCTTCTGCTAGTATGGGAAAGATGACCTGAAA  
ATCTAACCATTATTTTTTCCCCATTAATATCATTTTTATGATTATTCAGA  
AGTTAAATAATTGTCTGCTGTCTCCAAAAAGACTGAATCAACTAGCAA  
CAAATAAGAATTTTCTCACAGCTCTGCCAGCATTTTTAAAAGAATAGCTTT  
ATTGAGCCCAGGAGGTCAAGGCTGCAGTGAGCTGTGATTACACCACTCTA  
CCCGAGCCTGGGTGACAGAGCAAAACCCTGTCTCAAAAAAGAAATTTAAG  
GAACAGCTTTTATTGTTGTAATAATAGACATACAATAAACAGAGCACATATT  
TAAATTGTGCAACTTATACTTTGATATAACCCTGTGAAAACATCACCACA  
ATCAAGATAGTGAATATATTTATCACCTCCTGATACAGTTTAGCTCTGTG  
TCCCCACCTAAGTCTCATGTTGAATTGTAATCCCCAATGCTGGGGGAGGG  
GCTTTGTGGGAGGTGATTGAATTGTGGGGGTGCACTTCCCCCTTGCTGTT  
CTTGAGATAGTGAATGAGCTCTCATGAGCTCCCCCTTCACTCACTCTCTTT  
CCTGCTGCCATGTGAGGATGTGCTTGCCTCTTCTTTGCCCTTCTGCCATG  
ATGTGTTTTCTGAGTCCTCCCTAACCATGCCTCCTGTACAGCTTGAGAA  
CTGTGAGTCAGTTAAATCTCTTTTCTTATAAATTACCCAGTCTCAGGTG  
GCTCTTTATAGCAGTGTGAAAAGGAACATAATACCTCCTAAGTTACCTC  
AAGCTTGTTTTTAAATTCCTTCTCCTCCCTTCTTCAATGCCAAGCAAACA  
ACCACCTGTTTTCTGTCACTATAGATTAGTTTACATTTTGTGGGTTTTTT  
TTTTTTTTGAGACAAGGTCTGACTCTGTTGCACAGGAGCAGAGCAGCGTA  
TC

>Contig17

CGCGTTATAGGAGATGCCAACTTAAGAAATGATGATAAGGAGACTTTATT  
AAATATAATTTTGAATTATTTTGCCATTACAGAAATCTAATTATTTAAA  
ATTCTATTCATAATTTTTAATCACTGTACTTCCCAAGCTTAGCTTAGAAT  
CCTTCTGTGCTGAGGATTAATTTTAAATTTGTCTTTTATAGGCCTTATCTA  
AAATCCAAGAATAATTGCCAGAATCAACCACCTTCTAAATCTGTAAGTAG  
AAATTAGTCTTTTTTAAAAATATGCATTTCATAAGTATGATTAGTAATAAAA  
ATAATAAAGATGTTAGCAACCTAAAGAACATGTATTTGAAAGGTATTTCT  
TACAGATATAAAAAACAGTTTGGTTTAAATAAGAGACAATCATTTTTTGAAA  
AGTATGACATTTTTTGAAAAGTAGTTTAGTTTTATTAACCAAGAAAAGCC  
TCAAGTGAACCTTTAGTCCTCTTGATAGCTAACATTTATTGAATGCTTACT  
GTGTGCCTGATACTTTTCTGACTTGCATTACCTCACTGAGTCCTCACAAT  
CTTATGAGGCTACTATTAGTAGCCCCACTTTACAGATGAGCAAACTAAGT  
CACAGAAAGGTTAAATAGGTCGTATAGCTATTAAGTGACAAAGCTGAGAG  
CCTGTGATCTTAACCACTTTGGTATGCTGCCATGAAGTTAAATAGCTCAG  
TAGTCATTAAAGAGAACAATTTGCATTGAACCTTCCAAGCCACTTAACAA  
GTATATGCTTCCCTAATCAATTTAATTTAGCTACATTAGATAGAATGGTAA  
AGGATCCTTAACCTTAAAGTTTAAATGGAAGAAATTAGCCCTCTGAAAGAG  
GCACAGATTATTCATCTGCAATAAAAAATCTCACCTTTAGTTTTTTAAAAC  
ATAGTTTTTATCTGTGTTCTGAAATGTAACATAAACAGTGCTTCCTGAAG  
TGAAAAATTCTCACTGGTGAGAATTTAATAAGTTTTAATGATTCACCAA  
ATCACTTCAGTCATATTTAGTCATATGCATATGCATATATAGACATATA  
AGTTTTTATCTGTGTTCTGAAATGTAACATAAATAGTGCTTCCTGAAGTG  
AAAAATTCTCACTGGTGAGAATTTAATAAGTTTTAATGATTCACCAAAT  
CACTTCAGTCATATTTAGTCATATGCATATGCATATGTAGACATATATA  
TGTTGTATGTATACATGACATCATTAGACACTGTGAAGGATAGCAAAATG  
TATATAAGGCCAAAATTTATGAACAATGGTTTAAACGTTTGGGAAGCACTGG  
GTTACACTTTTACTTTTATGCAGATTGAACCAGTATAGTATGCAAGTCTTA  
AGGAAAAATCTACTGGAAAGGGCCCTCATTCACTTCCCAGAGGCTTCT  
CTGGAAGTTGACAATACTGACTTCAGTACATCAGCTCGTAAATGAGGATG  
ATACCTACCTTATCTGCTTTACACAGTTGTAAAAGTAAAAAGTGAACCTCA  
GGAAGGGAATTACAGAATTTAGGAGAACTAAAAGCACGATGTAAATAAT  
AGTCATCATTACAGTTATATAATGCTTGACAATTTATATAACACTTTTGA  
TACATGACAACAATAACTAACACCCAGACATGTTTATATACATTACCTCA  
CTCAGAACCAACCATGTGAGGAAGTTGGCCATATGCTTTAATGTCCAAACC  
AGGACACTTTTGAGAGTAAAAAGCAGTACTCTTTGACCAACAGGCATAAA  
TCAAAACTATCTTGTGAAAACCGGGATATATGGCATCCTTCCTAGATAAT  
AGATACTTTTACTATTATTAATTTTGCTGTGAATCTAACCTGCTCTAAA  
AAAGTTAATTTTAAAAAGTAATGAAGTACTGATACATGCTACAACATGGG

FIG. 3 (6 of 52)

8/118

TAAATCTTGAAAACGTTA TGCTAAGTG. .AGAAGCCAGACAGAAAAGGC  
ACATATTACATGATTCCATTTATATGACACATCTAAAATAGGCACATCTA  
TAGACATACAGAGACAGAAAGTAGACTAGCGGTTGCCAAGAACTGCAGGG  
AGCAGAAGATGGGGAGTGACTGCCAATANGAAAACGCATTACGT

>Contig18

TGAATCGCAATGATATGTGCCACTTTGCACTCTCTGTGACATATATAATT  
ATTTTTAATGCATTTCATTTTTTCTCAGAGTGCATTTCGTTTGAAAACATA  
GACGGGAAATACTGGTAGTCTTCCTTGTCAGTTAGAAACACCCAAACAAT  
GAAAAATGAAAAAGTTGCACAAATAGTCTCTAAAAACAATGAAACTATTG  
CCTGAGGAATTGAAGTTTAAAAAGAAGCACATAAGCAACAACAAGGATAA  
TCCTAGAAAACCAAGTTCCTGCTGACTGGGTGATTTCACTTCTCTTTGCTTC  
CTCATCTGGATTGGCATATTCCCTAATATCCCCTCCAGAACTATTTTCCCT  
GTTTGTACTAACTGTGTATATCATCTGTGTTTGTACATAGACATTAATC  
TGCACTTGTGATCATGGTTTTAGAAATCATCAAGCCTAGGTGAGCACCTT  
TTAGCTTCTTGAGCAATGTGAAATACAACCTTTATGAGGATCATCAAATAC  
GAATTCATCCTGAATGACGCCCTCAATCAAAGTATAATTCGAGCCAATGA  
TCAGTACCTCAGCGCTGCTGCATTACATAATCTGGATGAAGCAGGTACAT  
TAAATGGGCACCAGACATTTCTGTCTCCTCCCTCCTTTTCATTTACTTA  
TTTATTTATTTCAATCTTTCTGCTTGCAAAAAACATACCTCTTCAGAGTT  
CTGGGTTGCACAATTCTTCCAGAATAGCTTGAAACACAGCACCCCCATAA  
AAATCCCAAGCCAGGGCAGAAGGTTCAACTAAATCTGGAAGTTCCACAAG  
AGAGAAGTTTCTATCTTTGAGAGTAAAGGGTTGTGCACAAAGCTAGCTG  
ATGTACTACCTCTTTGGTTCTTTCAGACATTCTTACCCTCAATTTTAAAA  
CTGAGGAAACTGTCAGACATATTAAATGATTTACTCAGATTTACCCAGAA  
GCCAATGAAGAACAATCACTCTCCTTTAAAAAGTCTGTTGATCAAACCTCA  
CAAGTAACACCAAACCAGGAAGATCTTTATTATCTCTGATAACATATTTG  
TGAGGCAAAACCTCCAATAAGCTACAAATATGGCTTAAAGGATGAAGTTT  
AGTGTCCAAAAACTTTTATCACACACATCCAATTTTCATGGCGGACATGT  
TTTAGTTTCAACAGTATACATATTTTCAAAGGTCCAGAGAGGCAATTTTG  
CAATAACAAGCAAGACTTTTTCTGATTGGATGCACTTCAGCTAACATGC  
TTTCAACTCTACATTTACAAATTATTTTGTGTTCTATTTTCTACTTAAT  
ATTATTTCTGCAATTTTCCCAATATTGACATCGTGTATGTATTTGCCATT  
TTTAATATCACTAGACAATTCAATCAGGTTGCTACGTTGGTCCCTTGGGT  
TTACTCTAAATAGCTTGATTGCAAATATCTTTGTATATATTATTGTTTTT  
TCTCCTATCTTGTAATTTCTTTGAGCACATCCCAAAGAGGAATGCCTAGA  
TCAATGGGCACAAATAATTTGACAGCTCTTATTAAACATTATTCTGTAAAG  
TAAAAACTGAACTACTTTTCAGTATCACTAGCAACATATGAGTGTATCAG  
CTTCTTAAACCCCTCCATGTTAGGTCATTATGAACTTATGATCTAACAAA  
TTACAGGGTCTTATCCCACTAATGAAATTATAAGAGATTCAACACTTATT  
CAGCCCCGAAGGATTCATTCAACGTAGAAAATTCTAAGAACATTAACCAA  
GTATTTACCTGCCTAGTGAGTGTGGAAGACATTGTGAAGGACACAAAGAT  
GTATAGAATTCATTCTGACTTCCAGGTATTTACACCATAGGTGGGGAC  
CTAACTAC  
CATGCACACACAATCTACATCAACACTTGATTTTATACAAATACAATGAA  
TTTACTTTCTTTTTTGGTTCTTCTTTCCACAGTGAAATTTGACATGGGTG  
CTTATAAGTCATCAAAGGATGATGCTAAAAATTACCGTGATTCTAAGAATC  
TCAAAAACCTCAATTGTTTGTGACTGCGCAAGAAGAAAACCACCCATGCTG  
CTGAAAGTCAGTTGTCTTTGTCTCCAACCTTACTTCCTTTACCTCTCAT  
ATGTTTGTGAATAAGCCCAATAAGCAGACNCCTCTACAAAGTGAACCTG  
GTCTCTTTCTCCTAACAGGG

>Contig19

GTCTTGTAACACAGGTAAGACGAGTTCAAGTTTTATTTCTTGNTTTTAGA  
ACGGTAGGTGAGCGGTTTTGAGCAGTACACCTAAGGTAAGTAGCTG  
AATTGGGTTTTGTCTTGCTAAAGTTTAAACAACAGCTGGTCTTAATTT  
CTCCTTACCATTAGAGCACTCAGTAATCATATAAGTTGTGTGATCATTCA  
TTTTGCTTAACTGTTTGTCTTTCTGTTTTATTGCTGTTTCAGTCTTTTTCC  
CATTGGGTTTTGACCTACTCTATCTGACTTGATCAAATCCAAAGGAAATTT  
CCAAATTATGGGGAATGAGGCCTCTGAAGTGGCTAAATTCCCACCCTCCC  
ACACACACAAACGTGGTATGGTGGGGGAAAAAACGGCCAGCAAAAGAAAA  
AAAAAAGGAAAAGATGTTTCATTTTGACCACCAAACGGGCTTTATTTAC

FIG. 3 (7 of 52)

9/118

ATAACAAGGCCACCTTT...GCTAGCCA...CCATACTGAAAGAGCAATGL...  
TGTTGCCCCATGCTGTGGGTTCCATAGCTAACGTTCTGCCTTTTTTTCCTA  
CCACGACAGCCTGGGTTTGGTTCCTAAATCAAGCCTTTTCTGGTTTGATA  
CTTGGAATGCTGAAATAGCAGCAATTTGTCTAGCTGAAATATCGTAAT  
AAGATTTTAAAAGATTATTTTAAAGGACCTCAATAGTTAAAAGTCAGCT  
TAATTTAAAAGCTAACATCCAAGATGTGTGCATGTGTATGTATGCGTCTTT  
GTATTTAAATAGCCCTCATGTTTTTTTTTTCTTTCTAGGAACTTGCCTT  
TTTTTGAGCAAAAGTTTTTTTCTTCTGTGACTGGATTCTGTTTTCTT  
CATTTACTTCTGCTGTCTCTCTTTCTCTTGACCGTCTGCTGCATGAGA  
GCCCTAAAATAGTTTATAATAGCCTGGGTTCTTAAAGAAAATGGAGAA  
GGTGCCAGGCTCCCTTTTAGGGAGAACTTCTATTTTTCTTATGGAATC  
CCTAGAGTGTAACAGACAAGTTCAATTCAGCTCTTAAACTGCTTGCGTT  
TGTGTTGTGTTACCTGATTTTTTTGACTATTATTTTTGACTAGCTATT  
GCAACAGAAGCTACTCTTGGGTTTTCAAGGAAGATTGTAGTTTAGACATG  
TAGAAATGTCCTTTAAAAAACAACCTTTTTTTTAAAGTGCAGTGTAA  
AAGCATCATATGGTCTAGCCTCCTAATAATTTTCCCTTTTGGAGACCAG  
GATTCAGGGTGGGCTCTGCCCAGAGCTCAGAGATCCAGTTAAAAGAGAGG  
TAGTCTCGGCCGGGCTAGAGGCCAGCCTGTAATCCAGCACTTTGGGA  
GGCCGAGGCGGGCGGATCACGAGGTGAGGATCGAGACCATCCTGGCCA  
ACATGGTGAAACCCCGTCTCTACTAAAAATACAAAATTAGCTGGGTGTG  
GTGGCAGGTGCCTGTAGTCCAGCCACTCGGGAGACTGAGGAAAGAGGAG  
AATCGTTTGAACCCGGGAGGCGGAGCTTGCAGTGAGACGAGATGGCGCCA  
CTGCACTCCAGCCTGGCGACAGTGAGACTCCGTCTCAAAAAAAAAAAGAT  
AGGTAGACTCGATGTTGTCTGACCCGAGCAAGTTAGAGCAACGCCACACT  
TTGAGACGAATTTAAGAGTCTTTATCAGCCGGCGACCAAGAGACGGCTA  
ACGCTCGAAATTCTCTCGGCCCTTGAAGGGGCTTGATTTCTTTTATG  
CTTTGGTTTAGGAAGGGGAGGGGAGCTCAGTTGCAACAATTCTACAGGAG  
TAAAAACATGCAAAGAAATTAAAAAGACAAGTGGTTACAGGGAACAAC  
AGTTCCAGGTGCAGGGGCTCTAAATCTATCATAAGATGTTAGGTATGGGG  
GCTCTGCCGGACACAACTCAAGGCTTTATGCTGTTATCTCTTGAGCGAA  
ATCCTGGGAACTTCGTACATTGCTTGCTTCAGTACCTTATCAGTTAATCG  
GACTCTTTGATATGTTGGGAGTCAGCGTACACAAGTTAACTCCTTGAGGA  
AGGGGGTGGGTAAAGGAGTCTTGATGTCTGGTAAATGAAGGAGCGAAATC  
GAGTTCCTCTGGCTTTCTCAGCTAAGGGAGAGCTTATTCATGTGGAAACA  
AGGCTAAGTGATTAAGGGAGAAAGGGAGAGTCTGAAAACAAGGTTAGGTA  
TTACAATGTCAATAAAATTGGTCTCCTTATACAGTCTATGGTAGATTTC  
TTTCCATCTTTAATCTCCCTCTAGCACCACCAGACTTTTTCTCTCTGTAC  
CTTGAGATGTAAATTTTGCTATCTGAATTTTCGTCTAAGAGTTGTTTCCT  
TTAATATGCAAATTTAGGGTTATTTAGCTGACAACTGCCAAAGTAGTGAA  
ACAAGTTATCAAGAACTTGAACGTCTAAGGTAGGAAAAAAAAAAGTCTTT  
ATGAATCTATAAGATGTACTTCTATTGGCATGCCTAATACGTCTATGTAT  
TTACGTGTTGTGTACACAGTTTTTCACTACTGAAAATATATAGAGGAGTT  
CTAATTAATTGACTTAAGACAATAAAAGCGCTTGAATCAAATACCTTATC  
AGGAAAAAGGAAAAGACAAGTCAAATGCTTGTTCAAGTCTATATAACTTA  
AGTAAATCTTTAATAAATAAGCTAGCTTTAACATTATTTGAAATGTCTT  
AAGAATTGCCAGCAGGTTCTGGGTTACAGAACTAGTGGGGGTGCAGTGGG  
GTGAGGGTTGGTGGGGTGGNGGTNNNACNNNNNCNNCCCCCCCCCCCCC  
CCCCCCCCCCCCCTCCCCCCCCCGCCCCGCGGGCCGCGCCCCCCCCCGC  
CCCCCGGCCCCGCCCCCGCGCCCCCACCCCCCCCCCCCCCCCCCGC  
GCCCCGCCCCCCCCCGCGCCCCCACCCCCCGCCCCCCCCGCCCCCCC  
CCCCCCCCCCCCACCCCCCACACCCGCCCCACAGCACCCCCCACCCGAC  
GCCCCGCCCCCCCCCCCCCGCAGCCGACCCCCCCCCCGCCCCCCCCG  
CCCCGACCCCCGACCCCCCCCCCGCGCCCCCGCCCCCCCCCCCCCGG  
GCCCCCCCCCGCGCGCGCGCCCCACCCCCCCCCCCCCAGCCCCGACC  
GCGCGCCCCCCCCACCCCCCCCCCAGCCCCCGCCCCCGCCCCGACCC  
>Contig20  
GGCAGTACGCTATAATTCCTCTTCACCTTACCTCATCTGTTCTCTGATG  
GATGTACTTTTTTTTTTAGTTTTCTAAATTCCTTTTCTTTGCTCTGGAG  
ATGGGTGATTGATGTAGTCTGGGTATTTGTTCCCTCCAAATCTCATGTTG  
AAATGTAATCCCCAGTGTTGGAGGTAGGGCCTGGTGGGAGGTGTTTGGAT

CATGGGGGCGAGATCCC. ATGAATAGC. GGTACTGTCCTCTCAATAG. 3  
AATGAGTTCTCTGAGATATGGTTGTTTAAAAGTGTGTGGCACTCCCCCA  
TTGCTCTCTTGTACTGCTTTTCGACATGTGACATCCCTGCTCCCCCTTCGC  
TCTCTGCCATGATTGAAAGTTTCCTAAGGCTTCGCCAAAAGCTGAGCAGA  
TGTGGGTGCCATGCTTGTACAGCCTGCAGAACTGTGAGCCAAAATAAACT  
TCATTTCCATATAAAATTACCCAGCCTCAGATATTTCTTTATAGCAACATA  
AGAGTGGCTTAATACAGGCTGGGCATGGTGGCTCACGCCTGTAATCCCAG  
CCTGTGTGGGAGGCTGAGGGGGGTGGAACATGAGGTCAGGAGATTGAGACC  
ACCGGCTAACACGGTGAAACTCCATCTCTACTAAAAATACAAAAAATTAG  
TCGGGCGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTCTGGAGGCTGAGG  
CAGGAGAATGGCATGAACCCGGAAGCGGAGCTTGCAGTGAGCCGAGATT  
GCACCACTGCACTCCAGCCTGGGCGACAAGAGTGAAACTCCATTTAAAAA  
GAAAAAACAAAATTTCAAACAGAACAAAATGAAAAAATACCAAGTGAAA  
GGCCCCATAAAAAACCCCTCTGGGGCCCATCCTCCACCCCCCTCAAGTGA  
AACCACATTTAAACAATTTGGTGCATATCTTCCAAACCTTTTGTGTACA  
CATATAAAAAACATACATGCTTTGATTGGCTCAGACTGTACATAGTGTT  
TTCCCTCTTGCATTTTACACTTAATATATCTTTGACATCTTTCTATGTCA  
GTGCATGTTGGCTCGATGATATTCTATCATTAAATACCCCTTCCAAAAATG  
GTAAATCATTTTAAAAAATCATTCACACAAGTACATATTTACAATTTTA  
AAAGAAAACAGAATCCCAAAACACAACGACAAACCTCTAAAAATAATCTC  
TATCTTTCCACCAGCATGGAACAGTTTATTCCTTTTTACATAAAACGAA  
TTATGTGATTGGAAGATTAACTCTAATCTACACATTTATATACAGAATG  
TTCTATTTGTTAAGCCTATCTGAAAATAAAAAATTCAGATGATTAATTCA  
CTTACACTTAGAAATTAAGTCAATATACTATGAATACACATTGTGATCAG  
TTATAATATGATGCTTCTTAGTCTAGGGTTTCAATTAAATAACAGTAAAA  
AAAATTGGATAAATAAGACAGCTAATACTGAAAAATCCAGAAATTCAA  
GATTATATTGCCAACTAAAACACTGCCATTTACATTTTTTTTTCTTACTT  
GGTAGCAAATGCTAATGGAATTCAATCCTGATTACTTAAAGTCAGTTCAC  
ATCACACATTCATCAGGATAATACGAACATAATATGCCTACTATAGCGT  
TAGATTAAGACATAAAATTTTTTGCTTGAAAGTAATGACTGCGTACCAC  
TTGAGACATTTGTCAACCACTTCAGCACATTGTTTACGAGTGACTGGATG  
TCCACAAGGAATAAAAAACGACAGCAATATTTCTATCCATACAGATTTTGC  
AAAGCTTCTCCTCTTGAGGTGTCTTAGCTGCTCTTCAGTACTAATCTCT  
TTCTGCAATGAAGTCTGACTTGATTCTGCTTGTGTACTGTCTTTCTGAGC  
CTTCACTGGATCTGCAATCAGAACCTCAAGTGATTTACAGTTGCTCCCAG  
ATGTCTGAATTTTTCTCCATTATTTCTTAATGTCTTGAACTGAAC  
CCCATTCAATAGCTTCTTGTACCATAGGATTATGGAAGATGGTATCAAT  
TTTTCTAGTTAGTGATGGCGTTTTTTCAGCAGTTCTTACCAGACACTCCT  
CAAGTGAATGGGATAAATGAATATTGTTTATATATTTTCGTGTCTTCTGT  
TCTAACAGATATTTACACCCTGGATGCCATTAAACATGTTGTCCCAAGGGT  
CTTNCTGGGCT

>Contig21

CTTTCTCCCTTTTTTACCCCCATTTTCGTAGGGATTTGGTTAAAACCCATG  
TAAAAAATCCAAACACCGGCGGGGAACGGGGGTTCAAGCTCGTATCCCCA  
CCACTTTGGGAACCCAAGGTGGCAGGATTGTGGAAGCCAGGCATTTGAG  
CCCACCCTTGGGAAAAAAGAGAACCCCATTTTTTTTGAACAAAAACC  
CCAACCCCTCCCAGGAAAGAAATAAGTATGGCTGGGTGAAGTCACCAAAG  
ATGGCCGACTGGCTGGTCAAGTAACTTTACCTGATGGTTTCGTAGAATATT  
TACCTTCAACCAGGTGGGAGAATTGCTTGAGCCAACCTCAGTGTGGATT  
CAGGAACCTGATTTAATTGGTATCGTGATTGTGGATTAGATTCTCAGGGA  
TGCAATCACTAAGTAAAAGTGATAATAGCTACTTTTAAGTAAAATAATGA  
ATGAATCAAACACTCTAAATCCATGGTGCTATGCTAAGCTCTTTCTGTAT  
TTTATCTCATTTGATATTACAAATATTTGATGTGTTAATAGTAATGACTA  
TCTCCATTTTTTACAAGTAAGGAACTGACATTGAGAGATTAAAAGACTAG  
CACAAATCACAAGTAAATGAGATTTGAATCCGGTCTTGATTCCAAACTC  
TACAGTATTCTAAATTCAAGGAGACTAAATTATAAGATGGAGAGCCAATT  
TTACTTTATAACAGGGTTAGAATGGCAGAAGAGACCTGACATTCACACCT  
CTAGCCAGTGCATCATCTTCTGTAGGCAAATATGCAGGAAATCTATAAT  
AAGAACGTCCTTTGGTGAAGGCCAGGTGCAGGGGCTTACACTTGTAATTC  
CAGCACTTTGGGAGGTCAAGGTGGGAGGGTCGCTTGATGACAGGAGTTTG

AGAACAGCCTGGGCAACA TAGTGAGAC...TGTCTCTACAAACAAAAACA  
ACACAAAACAACTTCAAGAAAACCTCTTTGGTATGGATCAGAACAAGATG  
AATTATCTATCTGATCCAAATGCTTAATGACATTAAGCCACAGTCCACTC  
ACTGCCACAATAGAGATATACCTGCCAATGCCACTCAGGTAATCCCATCA  
AAAGTGGTAATGAGGTCTGCAGCATGACTTGTCTTAGTGATCCCAGCCT  
GAGACCTTGAGATTGCAGCATTTTATTCTACATATGCACAAAACATCTGT  
TGAAAAATCTTCTAAATTGATGCAATACATTTCGTATCAAGAATACCTGTC  
TGTAATCTCCATAAACCCCTCTCCTTTCTGTTTTTAAAAAATAGTAACAGCA  
TTTCTCCTTACATGACAAAGAAATGACTTCACCATCTACGAAATAGTGAA  
TAGGAGCTGTGTGGAAGGAAATTAGCTCTACTTCTTGGTGGAGATGAGAA  
GGGAGTGTTCCTCTGAAAATCAAGGCTCTTGTCTAGCTAGGAGCCAAAGT  
CGTTTTTTAGAGTGTGGACAGTTGAGAAGATAAGACAGGGACCATCCACT  
CATGTTTTTCTTATTCCATAGGCCTCTCTCAATTGGGCAAAGCACTCCAG  
ACCTTTTGGAAAGGTGACACCAAAGGCAAGCACCTGCTTGGCAGGCCCT  
CAGCTTCTACGCAAGTATAAGTGAGTATATAAAATGGGGGTACTTGTGCT  
GTTGAGTACCTTATTTCCAAATGAGGCCTGCCGGTGTCCCTGTGGCTGTG  
AGAAGGCCTCTACTGGATAGGTGGAAGTTGTGTGTTCTCATCTTTTCTAA  
CCCTGGATTGACTTGCCCAAAAGGAAGCCATTATTAACACTATAATAAAA  
CCATCCTTAATCTGGGACTCTCTTCATGCAGTGGTTCTTAACCAGTGATA  
AACATGAGAGTTACTTTTGGAGCTTAAAAAAATTAAGATGCTCAAGGTCT  
ACCCAAACTGACTGAATCTCCAGAGGTGAGGCCAGGGATGTATACTTTT  
GAGCCAGACCTCAGTTTACCCTGCAGAGCTCATAAGGTTGCATAACACCC  
TTTGTGAGCCACTCTGATGAAAAGAAAAATTGGTGAGGAATAAGTTTTAG  
AGAAGAAGGAGCAAAGGTGTTCTTGGCCAGTGAGAGCCAATGACAGGGAA  
ATGCAAAACATGTATCCACAAGAAAGGTAAATTACCCTATAGAGCATTTT  
AGGATAAATGAACATCTCATGCCCTAGGGTTGAGAGAGGGTACAAAAAAA  
AAAAAAAAAAGACCACTCTGGATACACAACGCGATAAATGGAATAAAGAA  
TTTTTTCCTTGTAATTAAAAAATCCTTTGTTACTGAGGTATAATTTAA  
TCTATTTTATGTATAGTTCAATGAGGTGTTATAGATAATAAATTTTTTT  
GTAAATTATTATATTGTATATACTCATACATTCATTTTTTAAAGTCAGA  
AATGTATATAACCATTAACTTATAAATCATTTCAGTCATTTCAGAGATATA  
GATACACGAGCATATTTTATATCCACCACAATAATTATTACCATCTCAAC  
AATTCCATCACCCCTCAAATTTCAAGCGTAGGGGTTTTTAAATGTCAAAG  
GAGTCTACTCAGTGGGAAGAAAGTTAAGGAAAAAACCTTTGGGGCTTTGG  
GCTCCTTCCCCCTGGGGTTAAAAAGGCAGGAAATTGGGCTTACCCCCCT  
GAAATTGGGAAGTAAATTTTGGGAAGTTTAAAAAAAAAAAAA

>Contig22

TCAAGCAGCCTTCTCTCCTTGGCTTCCCAAATTGTTGGGATTACAGGCAT  
GAGTCAGGATTCTGGCTTAGTTTACATTTTCTAGAGTTTTGTATAAATG  
GAAACATACAGAATGTATTTTTTTCGGAGTGGGGGAGTGTCTTATTTCT  
TTTCTTTCCATTTTCCCCCCCCCNCCCCCCCCGAGACGGAGTCTCGCTCTG  
TCTGTTGCCCAGGCTGGAGTGCAGTGGTGCAGTCTCGGCTCACCGCAAGC  
TCCACCTCCCGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCT  
GGGATTACAGGCGCCCGCCACCACCTGGCTAATTTTTTTTGTATTTT  
GGTAGAGACGGGGTTTACCATGTTAGCCAGGATGGTCTCGATCTCCTGA  
CCTCGTGATCTGCCCGCTTCGGCCTCCCTAAGTGCTGGGATTACAGGCGT  
GAGCCACCGTGCCCGGCCCAAGTGTTTCTATTTCTTAACCAGCTTTCATG  
CAATCTTTTTTTTATTTTACCATCTCTGTGATCCCACTCCCAAAGGTACTA  
GATGTCGATTGGTCCTTAGGATCAGCTACCATTGCCCCAAGTCTTTCCA  
GCCTTCCAAAAATTTTTTCTTTTTTCTTAAAGATACTCCTGTGTGAGG  
CTCAGAACTCTTGAATTGCTACTGCAAATATGAATCGGTGATGTGAATG  
CCAGGGAATTGCCTGATTGATCAAAGAAATGTATCCCTTCTCCCTCACT  
CTTGCTGTCTTCTCATTTTGTATTTTCCCCATCCTTGTGGATTTCGTGAATTTA  
AATATCCCTTTAATGTTATAAATTTTAAATGGCGTTTGGCGAAAAGTACA  
GAATTAGGTGCAAGAGTGATAGCTGTTATTTTTTTTTTTGGCCTCTGAGA  
CTGTTTATATATGCAAGTTATTTAACAGAAAGTTCTGCAGTGACCTGAGA  
TGTCAGGGGGGTCTGATAGAGTACGTTTGAAGGCAGTTACTGGAAAAAA  
TAATGCCATTTCTGTTTGTACTTCGGTAAGTTTCAGATGACCCAATATAT  
TGTTTACATGTGGCATTTCAGTAAAAAAGTAGCTTCCCCCTCCTTTCTTCT  
TCCTTTTCTCCTTCTCTGCTTCTATAAAGCATCTGCTTTGGGAACTTCT

TAGGAGGAGAGCTTGCCACCCCGTGGC...ATGGAGAGGTCTTGCCAGAGA...  
AAAAGAGATGCTCCCACTCAATGCAGGATGGTGTGGAGGTAAATGGGGAT  
ACGTCTGGCATCACTCAGGAATGGGCCTTCCTGGCAGGGAAAAAAGGGA  
GGGGAAAGAGGAAGGGAATTCNNANATNAATTGCTGAATACGGGGATTCC  
ATGGCCTGGATCCAGGAAGAGAACTTTGGGAGGTGTGAACCTGGAAGGCA  
TCANCTGATGAGGAGCAGCCTGAACTCCGGGGAGGACCTGTTTTTGGTGG  
CCCGGAAAAAATGCCTTCCACACACAGGGAGGCCACCCGGCTGATGGGC  
TGGGGGTGGACGGACAGCCCTAGGACAGGCTTGGGAAACCAGGCTCAGG  
TAGGGCCTGCGAGGTTCTCGCTGCGTCTCTTTCCTTCTTGGTCTTAGAAA  
ATAGAATCCAAGGCCTCTTGAGAGTGGAAAGGTGGGTGGGAGGAGGGCAG  
ATGGGGCTTAGGCCCAGGACACCCGTAGAGCTACTGCCAGCTGTCTCTC  
AGGGACTCTGCTGAGGTCACTCCAAGGATCATTCTTAGCCTTGCTAGACA  
GTACTGACAGAGGGAACCGTAGTATCGCACCCACTTCCTTCTCTTCAAT  
GAAAGTTTAAAGGTCAACATTTCTCTGGCAAAGGAAGTTCCACAAATAT  
TCCATTTCCGTTCTTAGAAACAGCAAGGTATCAAGCAATTGCAAACTTCC  
TGTGCTGGGGAATTTCCCAAGGAAGTAGGGGCAGAGTTCTGGTGGAGACAA  
AGTGAATTCGAGTGATTAGTCAGTAGCAGTAGCAGTAGCAGTAGCAGTA  
GCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGCAGCAGAACCC  
AGAATTTCCCGCACGTGTCTCAGGCTCTCATTGCGCAACTCAGTCTCTA  
AGTATTTTATTGGCAGGAAAAATAAAATAGCTATGAGTGAAATAATTCA  
TTAGACCTGAGCCTCCATCAATTTTGTGTTTAAAGGCCTGACTCTCTTTA  
CCTTCCCTGGGATGGAAGATGCAAATGTTCTGATCTCACTGTCAAAAA  
AGAAGAACCAGTGGGTATATTGTATGCTTGAGTCCAGCCATTAGTCACA  
AGACATAGAGATGACTGCCATGTGTGTAGACTTTCTATAGACTGTGTGCT  
AAACCCGACCTGCCACTTCCAAGGAGTAGATGAGGAATGTCCATGGTTCT  
GGGGAGCCCTACCCCAATTTGGGGCAGACATTCCAAAGCTCATTTTCTGT  
GGAGGGGGTGTGATGGTTAAAGGAACGGCTGGGATTTACTCTTCTTTCTAG  
GGCCAAGAAAATGACATGCTGCCTCCATGTTAATCATCCTTCCCCCTGT  
TAATAACTATGGCTTTAAGTCCCCGGTTAGGGCCTTCCTCCAAAATTGGG  
GAAAAAATTTCCCTCCCCCCTAAAAATTTTTTTTTTAAAAAACCTTT  
TTTTTGGGGGTGGGAAAAAACCAAAATTTTTTTTCCCAGGGGTTT  
TTTAATTTAAATTTCTCCCCAAAAATTTGTTTTTTTTTCCGCGAAAAA  
AAGACCCCCCAAAAAAAGTTTTTTGGCGGAAAAAATATTTTT  
TTTGTGTTAAGAAATGGAGAAGAAGGGGGTTTTTTTTTTCTTCTCCCC  
CACCCGCCAAAGGAAAGGTTGTTACAGATTGTTTTGTGTCTCCCGCCCA  
T

>Contig23

ATGTGCCTGCGAAATCATCCTTCCAGAAATATTTGCCCTTTCTTTTGT  
ATAGAGTGGCACTGCCCTATATGGTGACCACTTGCCACATGTGGCTGTTG  
AACACTTGAAATTGGCTTGTGAGAATTGCAGTGTAAGTGTAACACAT  
ACCAAATTTCAAAGACATGGCACATAATAAAAAATGTAAATATCTCATT  
ACAATTTTATATTGACTGTGTAAGTAACATTTTGAATATATTGGATTA  
AATACATGGATGATGCCCAACACCCACAGTCCCTTATCAAGTCTCTACT  
TCACATTTTGTACTTCTGACTTAGAAATAGCACTGGCGTCTAAGAGCCT  
ATTAATGTGCTCAATAGGTTCTTGGGAACCACAATTTTAAACAAAATGAC  
ATATAAGAAAACGAATAACATTGAACAAAATGACATTATTCGAGGACCTG  
CTGCATGTTGTTCACTTAAAGTCAGTGTCCAAGAACTATCAGTGACAT  
TTAGTGAGGAATTGCTGTCCTTCTGTTTACAGGAACCTGGGCAAGTTAC  
TTAATTCCTCTAAGCCCGGTTTATATCCCTGCAAAGAGAGAAGGATAATA  
ATCACCAGTACTTAGTGATGTGTAAGGAGAAAAATAAAATAATAATATG  
AAATGGCTGACAGTGTCTTGTACACAGAAGATGTGTGATCCACAGTAG  
CTGCTATTGCTGCTCCTCACTTCACTAGTAATGGTCCAGGGAGGCCTTTAA  
TGTGCATGGTGACGTACATTACATGTTGGACATGGGTGAAGGGAAAGAC  
CAGGCTCATCTAAACACAATAGGATGCTTGTGGTGTTTTGGAGGGAATC  
AAGGACTAGTTATCCACAGCTGTAACATGCATGGATCAAAAGAGATAAGG  
CACACAAAAGACTTTGTGAGTAGCAAAGCATTACAAAATGCAGAGACCAG  
CTGTGGGTGGTGGTGTGAGTCAGACCCAGCTTCCCTCTGTGCCTGGCTGAGT  
GGTTCTGGGCAAGTCACGCCATCTGTCTTGATGCCCTTCCCCATCTATAG  
AGAGGGAGCAACTGAGGCCCCCTCCAATACTGAAGTCCTTTATTTCTGCT  
ACTTTAGAAATATCCACATTTTGGTAAATTCAAATGATCCAATGATTCC

ATTTCTTAATGTTCAAAAAGAGCCCCAACATCTAAATGAATCAAC  
AATAAAATATTTATTGTGTATGTTTTGATTGCTGAACTTCTATTTTAGC  
AACACACACACACACACAGAACCCATAAGCCTTCATCTTTCCTTGGAT  
AAACGAGCCTTCTGTCTGGCCATTTAAGTCACGATTAAGTAAATGATTT  
CCAACCTCGCCTTTTGCAGCAGTTCAGATGGGTCTTTCCTGCGTGGCAGTG  
GCCCTCCTGACTTATGATTTCTGTGTGTGCGGCTGTTACCACTGCAGCT  
TAACTGAGGAAACAAGAACAAAACAGCCTCTGACCCCAAGAGACTGTTGG  
AGGCAAAGGCTTCAGTCCCAAGAACCTCACACGTGGGGAGCCCGAGAGCC  
CAGCCCTGACCTTTTTCTCCAGTAATAACATAAGAAACAACAGGCACTGGC  
CTTATTTTGGATACAAAGAGTGGTGCTTTTCTTAAATCTTCCTTTAGTC  
AGGGGTACCCCTTCATGGACGCCCCAACATCCATGGTTCTGCTTGAGTC  
CCTGCTTCCATATTCCTGCACTTCTCACTTGAAATATCCCTGGAGTACGT  
TAAGCAGCCAGGTTTGGAAAGTTCTTGCTGTGCAGGCGGGTGTGTGCATGT  
CCTCTCTCTCAACAGGACACAAGCTCCCCAAATCAGACGGTATGCCTCCA  
CGCCCCCTCCCAAGCCTCCCCAGCAGCACCAGCATGTGAGGGGAGCTGG  
GGCCAGGCCATGATGGGAAGCACTCTCTGCCTAAAGACTAGGGTGATGC  
GCCCTCAACTGTGGGAATGAGCCCCAGCTCTGGTGTCTGCCTCGGTTTTT  
CCTCCTGGACAATCAACATGAACTCCTCACCCCTCTTATCCACTTTGCAT  
AAACTGAAAATAACAAACCCAGGGTCTTTCTGTACAGGAAAGGGTTTTT  
TTTTATAAGATTAAACAGAGATGATTCAACACACCCAGGATATAACACAT  
GGGCCATGAGTCAAGGCCAGGCATTGCTCTGGTCAGCCTGTTGTTTGGGC  
CCCCCTGGCAGGGCTCTCCCTGAATCTTCCCCCTCTTGACTCCCCATCA  
CCACAGCACGTCCAGCTTTGGGTACAAGGCCAGTAAATGGGGAAGGGGT  
CAGATGACATAAAGAGCCCTTCTCTGTCCCATGAAATATATTTGGATAA  
CAGATGGCATTTCCTCTGTCTTGGCCAGGGCCAGAGCCTCCACTTG  
CTAGAGGCAGACAGAGGATGGAGAGCCCTTCATTAGTGGGAGGACATCA  
CAGGTGGGCAAGAAACCACAAGCTTGCACTGAGGCCAGCCTTGAAATAG  
CAGCACCTGCCGGCACCTGTGGTCTGGGGACAGGGTCACAGGATGGAGGG  
GCCTCCTAAGCCTTTTATCTCTATGTACTAAGTACAACCCATTTTCCAC  
CTCACAGAGCCAGATCAGCCTCTGTGAGGTCTGGTGGCAAAGGATAAT  
TGCTGCCCCGCTGCCCGCGGTGGGGTGCTTGTGCTTGCACTTCTGGGAA  
GGTGTGGGTACTCTGCAATAGGTCTCTCTGACCAGCTCACCTCCTA  
CTGCAAACCTCAAACCACTTCAAAGAAGATCCAGCACC

>Contig24

CGCGTAGTCTAAAGACTGAGTCTGAAGCTGTCCCTTCTGCTATGGACTT  
CAGATTTTAGCCCACTTGAATTGCTCCATATCCTCCAAGCCATGGCCATC  
CCTTGACTCTCTGGGCTCCCAAGCACTTGCTGCCTTCATCACACAGTTTG  
AGTTAAGGCAGAAAGACTGGTTTCCATGTACACTTTGTGGAAGCTTCTC  
ATTTCTTTATATAATCTCTGTCTTTGTCTACTGCTTTAAATCTAGAAA  
TTGTTTACAAACACAAAGGTGATCCTTTAAAGCTCAAAGCTGATTGTGT  
CACCAATATATACCACTCTTAATGGCTTCCCATTAACCTTTGAGTAAAGA  
CTTTATGGAGCCTACATAAGGCCATGACTACCTGGCTCTTATTTCTCTCC  
TCATCCTCATCTCACCACTCACTCTCCACTCCTATACCCCTCACTCCTT  
CCCCCTCCTCTCTGTAGCTCCAGACTCCCAATTACCTACTTCCACCCTT  
TTTGACCCCCAGGACTTATCTCAGCCTGGAATTTTCCCTCTTTGCTCTC  
CACTGAACTGTCCACTCCAGTCTAAGACATGTGCTTATGTCACACGCCC  
TTACCGTGCTTATCTCAGTTTGTAATTATCTACTCATTTAGAAAAGTGTT  
GATGAAGGTCTTCACTGTGAGCTTTCAGGATAGCAGGAATCATAGCTGAT  
TTTACTTACTTAACGGGGTTTCATTCTTTGTAACCTTTTTTTTTTTTGGAG  
ATGGAGACTCACTCTTGGCCAGGCTGGAGTGCAATGGCATGATCTCGGCT  
CACTGCAACCTCCACCTCCTGGGTCAAGTGATTCTCCTGCTTCAGCCTC  
CCGAGTAGTCTGGGATTACAGATGCCTGTACCACGCCCAGCTAATTTTTT  
GTATTTTGTGTAAGACGGGGTTTCATCATGTTGGCCAGGCTGGTCTCGA  
TCTCCTGACCTCAGGCGATCCACCCACCTCAGCCTCCCAAAGTGCTGTGA  
TTACAGGCATGAGCCACGGCACCCAGCCACTCCTTTTTTACTTATGGGTG  
AGAAGCCATTAGAGATCATTTCTTCTTTTCTTCTCTTCACTAAGGCA  
CCAGGGTCACTAAGTAGTAGGATACTTTGAACTAGAACTCAAGAAATTGA  
GTTTTAATTTTACCTCACACTCTCATATGAATTCTCCATGTGACCTCGGG  
CCATACTTCCCCTGTACCCTGTTTCTCTTTTATAAAAGTAAGAGTTTAA  
ACTAGATGGTCTCCGACATGCATCCTTCTCTAACATATTCTGGAACCTTC

FIG. 3 (12 of 52)

14/118

AATAAACTAAGATAAAAGGAATAATTAAAACCTTAATTTAAAAAGAAGA  
GGAAAGGAAGCAGTTACATTAAGCAAAAGAGACATCTTCATGGTTGAAGA  
AGTGTATGCCCTGGTGTCTGGATCCCATTTAGGAACTTGGTAACCTTGC  
AATCTTGGGCAGATTGCTTAATTTCTCTAGACCATGACTTCCTCTTCTGT  
AAGATGTGATAAGAACATCTACCTCACAGGTTTCATGAGAGGATTAAATG  
AGATAATGTATTATAATCCCTTGAACATGGTAGGCTGTTATGTTAAGTCC  
TTTCTCTCTTCTCTGTAGCTATCATGGAATTTAAAAACACATTATAACTA  
GAGCATGAGTTGCGACTAAAGGCTCAATTGTCTCTGCATGTGTTGGCTCA  
TGCATGCTTTTATTCCTCTGAAGAGCTTTTATACCAAGTGAAAGGAAATAA  
TTGCATTTCCCTGAAAATTACAGGAAAAAGTTATGTTTTTCTCTTCATT  
CAAGTGATTCTGTTAGACCCAACCATGCAACAATTTTAAAGTTGCTTC  
CAAATATATTTACAAATATTTCTGTCTTCAAGGAACAATGGCAAGACCA  
TGACTCAGGTTACATCCGGATTCCACCACTAACCATGTACCCAATTACT  
TCAGTCACCTTCATTGAGGTTTACATATCACAGAATAAAATCAGATTTT  
ATCAGAGGAGGTGAAGACAGGGAGAGATATTTCAATCCCTTCTCCGC  
AACCCCCGTTTTTTTTTTTTTTTTTAACAAGGATCCTAGAGTTACTGAATG  
ATAGCACGTTTGAGGGGAAAGACCCCTAAGGATGATCTTTATAAGCCATC  
ACTTGGTGTGTTGGTGATAAAAACTCGAGTATCTTTATGCAGTGGAAG  
GAGAAGATTGACTCGGAATCAGAAGCTTGAGTTCAAGCACTGGTTTCAT  
CAGTCTTGTGATCTTGGGTTGGTCACTTAACCTCTTCAAGGGTCTCAGC  
TGTGAAAGAAGATAGTATCAGCTAAFFETTTGTATGTGCAGTGAGGAGGCA  
GTGAGATAGTGCAGGTAACTATAAAACAATTGTCACATGAAACGCATCA  
CAGTGATTCTTTGGACCCACAAGCTCCAATCTTATAAAACATATCCAGTC  
ACCCACCAACATAGATCATCTCACCTTGCATATCTGATTTTGTGGATCAT  
GGGGAAAACTGCTGATTCTAGCAAAACCCATGGCATAGGATAAGTGCA  
CAATAATTTTTTTTTTCTAAATGATTTAGATGACAGTGAATTAAGGG  
TTTCTGAGGCCTCCTCAGAGTCGAGAGGTGGGTGCCTGAAGCCACCCAA  
AGTCCCTGTCACAGGATGGCTCCCAACGCACACACCACAGGCCTGCCCAG  
TATGTTCCACTATCTACCCAGTAGAGCCCTGCCAGTACGTTCCACTGTC  
CCTTCCCTAGAAGAGGTGACTGTTGTTACAGTCCCAGAAAAGCGGGCTC  
CCCAAAACAATGCAAGGACCCACCTCTCTCTGAACCTCACCCACCCTAGT  
TTTCTTTAAAAATCAATTTACAAGAAGATCATGTGAAGGAAAAGGTTGG  
GTGATATTCTAACCCAGTTAGCTGTTTCTCAACCAAGTTCTCTTTGAAA  
AATTCATCAACACCTTTGGGGAATTATTTACAACAGAGGAGTGAGGATG  
GGACCAGGATAGGTATTGCCTATGTTGGTGGAACCAGGGTTTTTTTTCTG  
GATTACCAAAGAGATGGTATGCATTGCTCCAGAAAGCTAAATATCTTCAG  
GCTTTCAATGGTGGCCTTCACCTGAAAATGTTATCCCTGTTGAAGCTTTC  
AAGCCAGTATTTTATAAGAACTATATTTTCTTGGTGAAGTGAAGGCATT  
ATAATGATGACTATACAGGTTCTTGAGTGAAGCCATCATTAGCATT  
GTCATTATTTTGTGTTAGTTGCATCTCCATAGCAGCTCACATTACAATG  
TGCTTTGCAATTGTTCTTAGCAATAGCCCTACAAGATTCTCAGGAGGA  
GAGGTTAATCCGGATTAACTTTCTGTGAAGCCTAGCGAGATTAATCGC

>Contig25

AAGAGTTTTAAATTAAGTAAGGACGCCGGGAAACAAATCAATCCCAGCA  
AACATTTTGTGGGATTATCATTCAAGCAATTTTACAGTTATCCCTGTC  
AAATACATTAAGTGTTCAAAATTGGGCATAGGGGGAACAAATAATAAAC  
CCAGCCAAAACAGAATAATCCCTGTTTGTTCATGTTGGATAAAAAAGAC  
ATTACTATTGGTGTAAGGAAATTAGATACATCTTCCATTATTTAGTAAAA  
TTACCATAACTTCTAACTTTGTGGCTTTAGGCAGTCTAGTCCACAGGCAG  
GAAGGAGGTTTGTGTTGGCAATGACTGTTATCATCTTCTGTTTCAAAGC  
TAAACCAATAACTAAGTTCCTCCCAAAGTTAATTCAGCATATGCCAGGA  
ATGAACAAGGACAGCCTGGACGTTAGAAGCAAATGGAGTCAGGTAGGTC  
AGATCTTCTTCACTGTCTCAGTGATGGCAGTTTCATAACTTTAAATGATG  
GCTATCACAGTTTTTATAAATAATCTAGATAAACAGTTAAATAAAATAA  
TTAGGTAAATGTAGTGCATAAATATTAGTAGACAACTCACCATAATTT  
AGAATCTAAAGTTAAATTAAATAATAATATTTTATTATTTGGTATTTTCC  
AAGAAAAACATATTGTAGGAAACCATTTCTTTTTAAAAAAAAGTGTCCT  
TTTAAAAAGGTGAATAATTTTGTCTAATTCAAAGTTTATTGAAAAGTTA  
TGATAAAACAAGGTAAAAGGAACAAGGAAATAAGGGAAATGTAAAGAAA

ATTATAGAAATAAAGTG...TATTTTTTGGTAAGAAAGCTTAAAGAGAA...  
ATTTTAGGTAAGAAAGAATCTTACCTAAAATTTTGTGCTAGAATAAAGTG  
ACTGGCTAAGAAAGGGATGTTCAAAGCTATTTATGACAAACCCACAGCCA  
ATATCATACTGAATGGGCAAAGCTGGAAACATTCCCTTTGAGAACTGGC  
ACAAGACAAGGATGTCCTCTCTCACCCTCTATTCAACATAGTATCGGA  
AGTTCTGGCCAGGGCAATCAAGCAAGAGAAAGAAATAAAGGGTATTCAA  
TAGGAAGAGAGGAAGTCAAATTTTCTCCGTTTGCAGATGCATGATTGCAT  
ATTTAGAAAACCCCATCATTTTCAGCCCCAAAACCTCCTTAAGCTGATAAGC  
AACTTCAGCAAAGTCTCAGGATACAAAATCAATGTGCAAAAATCACAGGC  
ATTCCTATACACCAATAATAGACTAACAGAGAGCCAAATCATGAGTGAAC  
TCCCATTACAATTGCTACAAAGAGAATAAAATACCTGGGAATACAACTT  
ACAATGGACATGAAAGACCTTTTCAGGGTGAAGTGCAAACCACTGCTCAA  
GGAAATAAGAGAGGAAACAAGCAAATGGAAAAACATTCCATGCTTATGGA  
TAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCCAAGTAATTTATA  
GATTCAATGCTATCCCCATCAAGCTACCATTGACTTTCTTCACAGAATTA  
GAAAAAATAAGTAGCCAAGACAATCCTAAGCAAAAAGAACAAAGCTGGAG  
GCATTGTGCTACCTGACTTCAAACCTATACTACAAGGCTGCAGTAACCAA  
ACAGCATGGTACTGGTACCAAAACAGATATATAGACCAAAAGAACAGAAC  
AGAGGCCTCAGATATAACACCACACATCTACAACCATCTGATCTTTGACA  
AACCTAACAAAAATAAGCAATGGGGAAAATAATTCCCTATTTAATAAATG  
ATGTTGGGAAAACCTGGTTAGCCATATGCTGAAAACCTGAAACTGGACCCCT  
TCCTTACAACCTTATACAAAAATCAACTCAAGATGGATTAAAGATTTAAAC  
ATGGCTGGGCATGGTGGCTCACGCTGTAATCCCAGCACTTTGGGAGGCC  
GAGATGGGTGGATCATGAGGTGAGGATGGAGACCATCCTGACTAACAC  
AGTGAAACCCCTGTCTCTACTAAAAAATACAAAAATTAGCTGGGCATGGT  
GGTGGGCGCTGTAGTCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGG  
TGTGAAACCAGGAGGTGGAGCTTGCAGGGAGTGGAGATCACGCCACTGCA  
CTCCAGCCTGGGCAACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAAA  
AAAAAAAAAGAGGATTTAAACATAAGACCTAAAACCATAAAAACCATAGAA  
GAAAACCTAGGCAATACCATTGAGGACATAGGCATGAGCAAAGACTTCAT  
GATTAGAACACCAAAAGCAATTGCAACAAAAGCCAATTGACAAATGGGAT  
CTAATTAAGCTGAAGAGCTTCTGCACAGCAAAAGAACTATTGTGAGAGT  
GAACAGGCAACCTACAGAATAGGAGAAAATTTTTTCAATCTATCCATCTG  
ACAAAGGGCTAATATCCAGAATCTACAAGGAATTTAAACAAATTTGCAAG  
AAAAAAAAAACCCATCAAAAGTGGGCAAAAGATATGAACAGACACATCTC  
AGAAGAAGACATTTATGTGCCCAACAAACATGAAAAAAGCTCATCATCA  
CTGGTCATTAGAGAAATGCAATTGAAACCACAATGAGATACCATCTCAT  
GCCAGTTAGAATGGCGATTATTAAGAGTCAAGGAAACACAGATGCTGGA  
GAGGATGTGGAGAAATAGGAATGCTTTTACACTGTTGGTGGGAGTGTGAG  
TTAGTTCAACCATTTGTGGAAGACAGTGTGGCAATTCCTCAAGGATCTGGA  
ACCAGAAATACCATTTGACCCAGCAATCCCATTACTGGGTATATACCTAA  
AGGATTAGAAATCATTCTATTGTAAAGACACATGCACATGTATGTTTATT  
GCAGCACTATTACAATAGCAAAGACTTGGGAACAACCCCTAATGCCACC  
AATGATAGACTGTGTAAAAAATGTGGACGTATACCCCATGGAATACTAT  
GCAGCCATAAAAAAGAAATGAGTTTATTCTTTTGCACGGAACCTGGATGAAG  
CTGGAAGCCATCATTCTCAGCAAACCTAACACAGGAACAGAAAACCAAACA  
CTGCATGTTCTCACTCATAAGTGGGAGTTGAACAATGAGAACACATGGAC  
ACAGGGAGGGGAATGTACACACCAGGGCCTGTCAGGAGGTGGGGGGCAA  
GGGGAGGGATAACATTAGGAAAAATACCTAATATAGATGACGGGTTAATG  
GGTGCAGCAAACCCATGGCACATGTACACCTACGTAATAAACCTCCAT  
GTTCTTCACATGTATCCAGAACGTAAAGTAAATTTAAAAAAGAAAGAA  
AGAAAGAAAAGGATGTTACAGCAAAACAGAAAGTCCAAGCATGTCATGA  
ATAGTCTGTGTAAGTCACAATAAGAGGATTTATTTAAAAAACTTTTATA  
TGATAAAGTTGTCTATAATTAAGGGAAATTATAATGGTCTTTCTAGAGA  
TTGGGTTGATGTTAAAAAATACTTATATATTAATAAATTGGTTAGAACA  
ATGAAATTTTCTTACGGGGTTGATTCACTCTTAATAAATTATAAGAGACT  
TAAGAATTTTTTTTAAACCAAAGTTCAGCTTTTATTGCATCTTGCTGTT  
TTAGGTTTCTCTCCCTTTTAAAGGGTGGGAAATAGTAATGCCCTCCTT  
CAACTCCCTTCAGCTCATATACGTTTTTTACCTCAGATTCTGTTTGTG  
TGTCCTGATGCTAACAAATGTTTTCTTAAAGGTCTAAAGGAAATGTTTTCT

FIG. 3 (14 of 52)

16/118

TCCAACATAATATTCTGTCATTGCAGAAGGTCTTTTCTTTTGCCTTTTG  
GTAAGTGGCTTAACAGATTTTATGTTTTATTGAAATAATTTCTATGCCAT  
TATTATTAAGTTTTGGTTTGCTTAGAAAACACTGAGATTAATACAATTTT  
TTAAAAATTATGATTATTACATCCATATATCTTTATGTATGTGCTTTTAA  
AGTCCTTGTGACATTGAGTTCTAGGGCTTGACTCCTGGGTCTTAAAGGA  
CAAGTCTGTCTAAATCTTAAATACTGACAGCAATTAAAGGCTCATCTTCA  
GGACTGGTAGAAAATGCCAATCAAAAATAAACTGCATTCTTGAAACACAGA  
GCCAGAAATTAAAGCTATTCAACTCAAGGCCAGGAAGTATAGTGGAAGA  
GGTGGGTGTGTGAGATTGTAAGGGCCAATTTTGAGAGATAAAAATAAGTTC  
AATTTCTCTATAAATTAATCATAATCATTGATGTCCAAGCCACACTGATG  
CAAGATCAGCATATGGGTCTGTGTCAGATTAAACAAGGTTTCTTGAAGC  
ATTAACCTACTCCTTAATAAAGGTTATAGAGGTTATAAAAGGCTTCTGGA  
AGTTATAGCTATGGTCAAGATAAAAAATTCATAGATTGTTAATACAATTT  
TGGAAAACAAATTTAATTGGCTTCTTGCTGTTTTTATTAGGGCTTATTGT  
TTGAAAATTAAGTCTCGTCTCTCAAAGAATGAAGGCTTTCACCTTTTTT  
TTTTTTTTTTTTTAATCCTTGAGTTATCACTTTGGTCAAATGAATGACTTA  
TTTTACAATGACCTTTTCATCAAGTGTTTTAAACCTTTCAAATTTGACAAA  
CTTTCCAAAATCAAATACTACAAATTATGTCTTTTTATGACCTAATGAATCC  
TTTAAATACTAGGTTCCCTAAAGTCCAAAAAATAACATAA  
TGTGGCTTATTTGGTATAAAAAATTTTACAAGAAACATTGTCAAATATAAA  
ATATTGTGTGGTTTTGTTTTGGGCTGTATTTGTATAAATATGTTATTGGTA  
TGTGTTCCAAAATTATAGGAACTCCTATAATTCTGATATGACTTGGTGT  
ACATTATCAGTAATAATTATAATTGTTATGGTAAATTATTGTGTGCCATG  
GAGGTAACAAATTTCTCATCAAGTGTGTCTTTGACTATGGTTGCCCTAA  
AATTTTTGCCATTACAGACAATTGTCTTGCTTTGGTCTCTTTAGAAG  
GTGGTTTTATGTAATCAGCTATAAACTCTAACGGGTGCTCTTGAATGCAGG  
CTTAAGATAGCTTTGGAGACTGTGACATCAGAATAGAGAAAACTTTCA  
GTATTATGAGGTGCTGAAATATTCATGAATATCAAGCAAAACAGGAATT  
AACTTCATAGATGGAATAAAAGAATGCTGAAGTAATCTTTTTGACTTTT  
TTTCTTAGAATGTTGATCCTTCGTTTTGTTTTTCAGAGTCNAGGAAATTT  
TTCTGTTGAGATATTGACAGCTTTAACAATTAAGTATACTCCAGTGAACA  
CAATTTGGAGCA

>Contig26

ATCTAGTCATTCCCCAGCCTGACCAATTCAATGGCCCCCATCTTAGTTAA  
AATTCCTCACCCTGACAAGGCCCATCTACGCCCTGACCTCATGCCCTC  
CACTCTCAGTCTTGCACTCACCCTGCCACACTCAAGGGCTTCCCCAGGTT  
CCTTCTTAGATTCCACCGATAGCTCAGGGACTTTGCACATGCTACGGTCT  
CTGCCCTGGCTCCTCCCCAGATCTTCTCATGCCCTAGCTGCTTCTCATCAGC  
ACCCCTCAGAGACTGTCCCTGCCCCACCTCTCCAGGTTCCATACCTGCCA  
CCCTCCCCCAATCAGTAACAGTTTTCTTACAGAGCGAGTTACCATCCCA  
GTATTTCCCTAACTTATTTTTTGTGACTGGTCTGTTGCCCTGTCTCCACCA  
CAAGAACATAAGCTGCATGTGAACAGGAGCCTTGCTATCTTGTCAACCC  
AGTGGCTGTGACATAACCTGATACACATTAGATGCTCAATGATGTTTGAT  
GAATGAAGTGCTGGTAGTCCAACCTGTGTTTCTTGTCTGTGTAAGTATGT  
CTGTTGTGGTTTTCTAAGAACCCTACAGCTCTCCACTGTGACTCCTGTTT  
TATGGTCTGATTTGCTGGACTAGAATCCTAACCTACATGCTTACTCTTA  
GTGTCCTCCCCAGAGGCTGAATCCCAGTCCCTAAACCTCCACCAATGG  
CTAAGACCTAGCTTCCAACAGACAGGCTACGCTGAGACCTCAGCACCG  
CCCTTCTGCGGTCTCATCCTTAACGCATCCTTCAGGGCCAGCTTAAATG  
TCTCTTCTCCAAGGAAGGCTATCCTCTTTCTGCCCCCTCAGTGCTCTCCAT  
GCCTCCTCTATGCCTCCATGCCTGCTTTCAACCTGCAGAAGTGGAGAAA  
TTGCTAATCTGCTGTGTTGACACTGTGCTGGGGTGCCTTGGGCCAGGGAG  
CAGGCTGGTGGTGTGCTGATAGCCCGTGGCTGTGCCAGGTCCATGCTCA  
CTTCTGAGCCCCAGTGGAGTAGGCTCCCTTTCCCTTATTGCAGCACTCA  
GAGGAAGGACGTGCTTCTTAGGACAGATCTGGCCAACCTCTCCCTCGTGA  
GAGAAGGCCAGCCATCCTCTTGCCCTCTTTCTTTCTCCTGCCCCCGAGT  
AATAAAGGTGCCTGGTCAGAGCCTTCTAGAAGGAGACCCAAACATCCACC  
ACACATTCCAGTTCCAACCGTCATCCACATGGCTGGCTGTGCAGGTAAA  
CGCAGAGTCTGTTTCAACACCCCAACCATCTAGTATTGGATGGGAGGACA  
GTAGCGTGACACTCTTCTCCAGCCTTGAGCCCTACTGTGGGCCCCACCCA

>Contig27

**FIG. 3 (16 of 52)**

18/118

CACACACACACAACAGCTTAGATGTTTTCTCCACTGATAAGTAGGTG  
ACTCAATTTGTAAGTATATAATCCAAGACCTTCTATTCCCAAGTAGAATT  
TATGTGCCTGCCTGTGCTTTTCTACCTGGATCAAGTGATGTCTACAGAGT  
AGGGCAGTAGCTTCATTCACTGAACCTATTCAACAAGCATTATTCACTGAG  
AGCCTTGTATTTTTCAGGCATAGTGCCAACAGCAGTGTGGACAGTGGTGC  
ATCAAAGCCTCTAGTCTCATAGAACTTAGTCTTCTGGAGGATATGGAAAA  
CAGACAACCCAAACAACCAAAAAGAGCAAGATGCTGCAAAAAAAAAAAAA  
AAATGAATAGGGTGTCTAAGATAGAGAAAAAGTGGGAGAGTGCTATTTAGAC  
AAAGTGGTAAAAACAAAGCCCCCTTGTGAGATGAGAGCTGCCGACAGGAGG  
GGGCGGGTCATGGTTGTGGGTTTTTGGGTAGGACATTCAAGAGGAGGGGGC  
GGGTCGTGGTTGTGGGTTTTTGGGTAGGACATTCAAGAGGAGGGGGCGGGT  
CGTGGTTGTGGGTTTTTGGGTAGGACATTCAAGAGGAGGGGGCGGGTCTGTG  
GTTGTGGGTTTTTGGGACATTCAAAAGAGTCTGAATGCACCCAGGCCTAC  
AACTTCAAGATGGTAAAGGACAGCTCCAAGGATCAGAAGAAGCATGCTTG  
GAACTGGGGCATTTTGAGAAGGAGGAAAAATATGCAGAGACTAGTGCTTG  
CAGAGCTTGCATGTGGATTTCAATTTGAGGTACAATGAAAACCCATTAATG  
GGTTTCACACAGTGCAATGGCCTGACCTCACTTATATTTCTTAAATAGA  
AAACAGATCAGGAAGGAAGCAATAGAGAAGCAGAAAGTCCAATGAGGAGG  
TTTCACAGCAGTCATGGGGGTGGGGTAAGGAAAAGAAGTGAAAGAAACA  
GACAGAATTGGGTTATATTTTGGAGATAGAACCAACAGAAGGAAGAGGAG  
AAACAACATTTACTGAGAAGGGAAAAAGTAGGAGAGGAATAGGTTTGGGA  
AATAAATCCTGCTGACATTGGAAACCCCAAGGAAGCCTCAAAGTATATT  
TACTTGGCTTTAGATTTAAAAGAATAGGAAAGAAGCATCTCAACTTGGAAT  
TTGAAATCTATTTTTCCATAAAAGTATTGTTAAATTCTACTCATACTCAC  
AAGAAAAGTACATTCTAAAGAGTATATTGAAAGAGTTTACTGATATACTT  
AGGAATTTTGTGTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTTAAC  
CTTCAATTGTTGACTTAAATACTGAGATAAATGTCATCTAAATGCTAAAT  
TGATTTCCCAAAGGTATGATTTGTTCACTTGGAGATCAAATGTTTAGGG  
GGCTTAGAATCACTGTAGTGCTCAGATTTGATGCAAAATGTCTTAGGCCT  
ATGTTGAAGGCAGGACAGAAACAATGTTTCCCTCCTACCTGCCTGGATAC  
AGTAAGATACTAGTGTCACTGACAATCTTCATAACTAATTTAGATCTCTC  
TCCAATCAACTAAGGAAATCAACTCTTATTAATAGACTGGGCCACACATC  
TACTAGGCATGTAATAAATGCTTGCTGAATGAACAAATGAATGAAGAGCC  
TATAGCATCATGTTACAGCCATAGTCCTAAAGTGCTGTTTCTCATGAAGG  
CCAAATGCTAAGGGATTGAGCTTCAGTCCTTTTCTAACATCTTGTTCTC  
TAACAGAATTTCTTTCTTTCTTCATAGGAGATGCCTGAGATACCCAAAA  
CCATCACAGGTAGTGAGACCAACCTCCTCTTCTTCTGGGAAACTCACGGC  
ACTAAGAATATTTACATCAGTTGCCCATCCAACTTGTTTATTGCCAC  
AAAGCAAGACTACTGGGTGTGCTTGGCAGGGGGGCCACCCTCTATCACTG  
ACTTTCAAGATACTGGAAAACCAGGCGTAGGTCTGGAGTCTCACTTGTCTC  
ACTTGTGCAGTGTGACAGTTCATATGTACCATGTACATGAAGAAGCTAA  
ATCCTTTACTGTTAGTCATTTGCTGAGCATGTANTGAGCCTTGTAATTCT  
AAATGAATGTTTACACTCTTTGTAAGAGTGGAACCAACACTAACATATAA  
TGTGTTATTTAAAGAACACCCTATATTTTGCATAGTACCAATCATTTTA  
ATTATTATTCTTCATAACAATTTTAGGAGGACCAGAGCTACTGACTATGG  
CTACCAAAAAGACTTACCCATATTACAGATGGGCAAATTAAGGCATAAG  
AAAATAAGAAATATGCACAATAGCAGTTGAAACAAGAAGCCACAGACCT  
AGGATTTTCATGATTTCAATTTCACTGTTTGCCTTCTACTTTTAAGTTGCT  
GATGAACCTCTTAATCAAATAGCATAAGTTTCTGGGACCTCAGTTTTATCA  
TTTTCAAATGGAGGGAATAATACCTAAGCCTTCTGCGCAACAGTTTTT  
TTATGCTAATCAGGGAGGTCATTTTGGTAAAAATACTTCTTGAAGCCGAGC  
CTCAAGATGAAGGCAAAGCACGAAATGTTATTTTTTAATTATTATTATA  
TATGTATTTATAAATATATTTAAGATAATTATAATATACTATATTTATGG  
GAACCCCTTCATCCTCTGAGTGTGACCAGGCATCCTCCACAATAGCAGAC  
AGTGTTTTCTGGGATAAGTAAGTTTGAATTCATTAAACAGGGCATTGTG  
GTCCAAGTTGTGCTTATCCCATAGCCAGGAACTCTGCATTCTAGTACTT  
GGGAGACCTGTAATCATATAATAAATGTACATTAATTACCTTGAGCCAGT  
AATTGGTCCGATCTTTGACTCTTTTGCCATTAACTTACCTGGGCATTCT  
TGTTTTCAATTCACCTGCAATCAAGTCCTACAAGCTAAAATTAGAT  
GAACTCACTTTGACAACCATGAGACCACTGTTATCAAACCTTTCTTTTC

FIG. 3 (17 of 52)

19/11/8

TGGAATGTAATCAATG1 . FCTTCTAGGTTCTAAAAATTGTGATCAGACCA  
TAATGTTACATTATTATCAACAATAGTGATTGATAGAGTGTTATCAGTCA  
TAACTAAATAAAGCTTGCAACAAAATTCTCTGACACATAGTTATTCATTG  
CCTTAATCATTATTTTACTGTCATGGTAATTAGGGACAAATGGTAAATGTT  
TACATAAATAATTGTATTTAGTGTTACTTTATAAAATCAAACCAAGATTT  
TATATTTTTTTCTCCTCTTTGTTAGCTGCCAGTATGCATAAATGGCATT  
AGAATGATAATATTTCCGGGTTCACTTAAAGCTCACATTACACATACACA  
AAACATGTGTTCCCATCTTTATACAACTCACACATACAGAGCTACATTA  
AAAACAATAATAGGCCAGGCACGGTGGCTCAGACCTGTAATCCAGCAC  
TTTGGGAGGCCAAGGTGGGAAGATCACTTGAGGTGAGGAGTTCAAGACCA  
GCCTAGGCAACATAGTGAGATCTCATCTCTACAAAAAATGAAAAAT  
TAAAAAATGAGCTGGACATGGTAGTACACACCTGTAGTCCCAGCTACTCG  
GGAGGCTTGAGGTGGGAGGATCACTTGAGCCTGGGAGATGGAGGCTGCAG  
TGAGCCATAATCACACCATTGCACCCCAACCTGGGCAACAGAGTGAGACC  
CAGTCTCAAAAGATAAATTTTTAAAAATGTTAAAAATATATAAAAGAGA  
ATTTTAAAGAACAATAATAGATCAAAGCATGGATGCAAGATATATTTA  
GTTGGAAAATCAAGGTTAAAATCAAGGGATCTTGGAATTAGGTGTGGTAG  
ATTTGGGTAAGGAGTAGTCTAAGATGACCCTGTTTCTTGGTACTGGAGAC  
TGGATGAGTGCCAGCGTCTTAACCATATTTTTGGTAGAAATATGGAGGTC  
TTCTCCATTCCAGGATGAATGATGAGTAAAATTTTAGGCATGTAATTTGA  
GCTACTAGAAGGACACTCAATTGCAGATGTACAATGGGGAGATGATAACC  
TATCTGGAACCTCAGAAAAATAACTGTATATAGATATGAAAGACATCAGTA  
GGTATGTAGTAGATAAAATCCTAAAAGTGATGTCAAAGGGAGAAGAGAAG  
TATATGGTGAACACTGTTGTTTGTCCATGCAATTGCCATCTCTTCTTCTT  
CCTTACTGACAGAACCCCTGATTTCACTGAGAAGTCAACATGCCCTTCCCC  
AATTGATGAATCCAATTGGTTGAAGATTATGTTTATTCTATTCTTACATG  
ACTAAGTCAAGTTGACTTAATCCTATCAAATGAGATGTGATCTGGAAAC  
AACTTCTGGAAAAGATTTTCTACCTTGATAAAATAAAGAGCCATATAGAT  
GGTCTTTTATCTTCTTCTTCTTGAATGAGATATGTTCTATGAGGAAGT  
GAAGCTTAGAAGTGTGGTCAGCAACTTGCAACGACTGGGAAGTCAGAGCC  
ACACAATGAAGAATGCAGAGTGAAGGAGAAAAAGAGCCAGCATCTCTGA  
CAACATTGTTACACCGAGAACCTACCTCCAGATTTTAAAGAAAACAAGAAA  
TGCTACTGTTATTAAGCCATTTCACTGGGTTTGCTATGACTTGCACTCAA  
ATCTAGCTTAACTGATACAGAGCACCACAGAGAACTGGTCTCTCATTGT  
CTCATCTGTTCTTTCTAGCAGCCACGACTTTCTAGGGTTTCTTAGCC  
CAAGTCTGGCTAGAGCAAGACTAAGTAAGACTTGATTCTTAAATGTCCTT  
TTGTTTTAAGAAATATTAAAGAATTATTTTTATATTAATATATTTTAAGA  
AATAAGGAAATACAAAACACTGAGCAAGCAACACAAATTCAAGAAATCTT  
AAAAAGTATAATAGCTGCTCAGTCTCTGATTAACAGTGAAATATGGAATC  
ATTGTAGAAATGGCCTTGAGCGTTATTCTCCAGGCCAGCTATCCTTAT  
GGTCTGCCCCACCTCCCTCATTCCTAAACAGTAAGAGAGTCCCATGGTG  
AGACTCAACAGTCTTAGCACAGAACTTGTTACAGTCTATTTCTTTCTTA  
CAGTCTATATATCAATTCCAAATCAATGAGAGTAAAGCCCAATCCCTGC  
CTTTAAACCCAAAGGACAGAAGCCCAAGCCCAAGATATTCCCTAACCT  
TCTCCCCCT

>Contig28

CCTGTCGCTCCCTATGTTTAAAGCTGGGGATCTCTTTTTCTGTGTCTAA  
TTATTTTCTCATTGGCTTGAAAAATCTGATAAAACATTTTAGGACTGTG  
TATAAAATAGAATTAGCCAAGTGCAATGTCTTTATTGAGAAGAAATTCA  
TGGACGTTGTGCCTACTCTCTGGCTTCTGGCTTCATGGCTTTCCAGAT  
CCCACAGTAAGCTCTGGATAGTAGAAGTTATAGTAAGACTGACTTCTAAA  
TAAATGAAGTGACTTTAACCTTACTGATATGGCTTAAAGAAAAGGAGTGG  
CCTTTAAGATCCATGAACTTCTCAAACAAAAGTGATAACGTTATCTCCAT  
GCATATATAATACTAAATATAATGCAACTGAGAGAAGTAGGCTGTGGTAA  
GAAAGGAGACCCAAGTGCCATCTGAAGGCAGCACTTACCACTCTGCTTCA  
TCCACCGAGGAAACAAAGCATGAGTATTGCCAGATTTTCTTCTGTTTCA  
AGAAAAGCCAGAAATCCAGGTTTTTGCCTGAAATGTCCTGATTTAATGT  
TGGGAACATAATTTATATTTTGAATAACATTGTGTGGGACAAGTGAACCT  
GTATGTGGAAGTCTTTCTCCAGTGGCGACCACTTTGGACCGTTGATAC  
TCAGCAAGTTGAGCCAAGTGCGCCTTGTCATTGTGAGTCATCAAGGTGAT

STGTGATTGGTCAAACAATTAGTTTTGCTCAGCATCTCGTGTGTTTTCA  
AGGACCTGAGGGTTCATTTGCCCATGCAGATCTTGTAGTCTGTATTTC  
TATTAATTTATCTTGCAAATCTATAATGTTTTATTTTAAGCAGCGAGAGC  
CGTGGCAGCCTTTGGTCTGGACCTTTCTAATGATCATTTAGTATCAGGC  
TATGTGGGAGTTGATTGTTTTGCATTGCCTGAAAGCCAACAGTATCACTC  
CTCCTCTAGGTGTGGCAGAGATGTGAGAGAGGGAGACTGACAGTCTGTGG  
GTGTGTATGCAGTGTGGGGGAAGCGAGGCACAGGGGACAATACTGTGGT  
GTATAAAACTAGTCTAAGGTAGCATCAGGAAGTTCATGAAGCCAAATGA  
TTTTATAACAGCACAAAGACATTATTGTTTTTGCCTCCCTCTCATTTTT  
TTTTTTTTTTGAGACAGAGTCTTGCTCTGTCTCATCCATGCTCGTGTGCAGT  
GGTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGGGTTCAAGCAATTC  
TCATGCCTCAGCCTCCTGAGTAGCTGATTACAGGTCTGCACCACCCCGCC  
GGCTAGTTTTTTGTATTTTTTAGTAGAGATGGGGTTTTGTAATGTTGGCCAG  
GCTGCCCTGTCATTTTTTTTTACTAGTGTCCAGTGGAGTTTTTTAGGGG  
CTACATAACATGATACTGTCAATTAATCTAATGGCTAATGAAAGGGATATG  
TATATGTTTTTTGTGTTTAAACAAACTTCTTTGGGGTCTCAATAATTTT  
TAAGAGTATAAAGGGGTCTGAGATCAAAGAGTTTGAGTTCTGCTGGACT  
GGGACAGTGGTTGTCAACCCAGATTGTACATTAGGGTCATCTGGGAAGCT  
TTAAATAGTACTGATGCCCAACCTTACCGCAAACCAATTAAGCCAGAAT  
CTCTGTGGATGAGAAGTCTTCATTGTCTCATCATCATATCATCTTCATTATC  
TGTCACCGTCACTACACCATTAATCATCATCATATCATCTTCATTATC  
ATTGTTAGTATCTCCATCACCATCATCAGCATCACCATTATTATCATCAT  
CATCATCCCCACCATCATCCTCATCGGAACCTCACCTGCATGGAGGACAA  
TCCACTATGCATTAGGTGCTATGCTATTTGCTATACTCCTTATTCTCACA  
ACTGCCCAGAGAGGGCTGATATTATCTCACTTTATAACAGGAGGAATCTGG  
ATCGGAAAAGTTAAGGTAAGCTAATTCACAGAGCGAGAAGAGATAGAGCC  
AGGATTCGAAACCAGTTCTCTGCTACATCAATGTTCCAGTCTTGCACT  
ATTGAGAACCTCTTTAGTTATGCTTTTACCCCTCCAACACCACAGTAAAT  
TTTTCTTTTTTTTAAAAAAATTAATCTTTAAGTTATAGGGTATATGTGCA  
TAATGTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTGTGCTG  
CACTCATTAACCTCGTCATTTACATTAGGTATATCTTCTAATGCTATCCCT  
CCCCGCTCTCCCCACCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACC  
CTGTGTCCAAGTGTCTCATTTGTTTCACTTCCACCTATGAGTGAGAACAT  
GTGGTGTGTTGTTTTCTGTCTTGTGATAGTTTGCTCAGAATGATGGTTT  
CCAGCTTCATCCACGTCCCTACAAAGGATATGAACTCATCCTTTTTTATG  
GCTGCATAGTATTCCATGGTGTATGTGTGCCACATTTCTTAATCCAGTC  
TATCATGTCTGGACATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATA  
GTGCCACAGTGAACATTCATGTGCATGTGTCTTTATAGCAGCATGATTTA  
TAATCCTTTTGGGTATATACCCAGTAATGGGATGGCTGGGTCAAATGGTAT  
TTCTAGTTCTAGATCCTTGAGGAATTGCCACACTGTCTACCACAATGGTT  
GAATTAGTTTATAGCCCCACCAACAGTGTAAGCAATTCCTATTTCTCCA  
CATCCTCTCCAGCACCTGTGTTTCGTGACTTTTTAGTGATTGCCATTCT  
AACTGGCACCACAGTAAATTTTATAGATTTTATAAGCAAATTGTATTTA  
CTGTGCAAGAATTGGTTTATTTTTTAAACCATGTGTTGCAAACATACAAT  
GGTTAATTGTGATATTTGCTCAGTACAAGATCATCAGATCACTACACAGA  
CTTGAGGTAATTCACCTAAAAGCAAAGAGAACTGACCCACATTAAGT  
AGAAGTCTTTACTTATTTATTCCTATAAACGAGCCAATATGAAGAGAAG  
GCCTTAATGTGGTTAACTATGTAATTTTTTTCTGACTTTTTGAAATACTG  
AGAAGAGCTCATGACTCTCCCATCTCCTAATTCTACCTTGGTGGATTTTA  
GACTGACCACAACCTCATGGGTAAATGAGGGAAGACGAATAAGAAACCTTG  
CTTTTTTTCTCTCTTGTTTTTTGGCTGGCTGCAGTGGCTCACACCTGTAA  
TCTCATCACTTTGGGAGGCCAAGGTGGGAAGATCACTTGAGCTCAGGATT  
TCAAACTGGCCTGGGCAACATAGTGAGACCCCATCTCTAAAAA  
AAAAAAAAAAGGCGACAGGCGGTGCGTGCCTGTAATCCTACTACTC  
AAGAAGCCGAGGTGGAAGATCACTTGAGCATGGGAGGTCAAAGCTGCAG  
TGAACCTTGATTGCACCACTTCATTCCAGCCTGGGTGACAAAGCAGGACG  
TGCCCTCAAGAAAAACAAAACCTTAATTTTTTGGCTATTCTTTTC  
TGGTAAGAATGGTATAGAGATGGGGATGAGGATGGCTATTGTATGAGAGA  
GCAACAGGGTCCAAGCAGTGTCTGGGCTGTCTAAGGACCAGTAGTCAG  
CTTAACCTCTCAAATTTCCAGGGAAGGAGTTCGAGTGGTAGAATATCCT

FIG. 3 (19 of 52)

21/118

GGGTATGCCCAAAGCATCACCTTGCAAATAGCCTGTGCATGAATAATTTG  
TTCATTTGTTATGACTGGAACTGGCTTTGTGTATGCCAGAGAATGGGGG  
CAGGAAAGAGAGATTGGTGTCTTGAGCTCTCTGTGCCTCTGGGGCAGTGA  
TGCTTTTCTCTCATGTGGAAGGAGAGCATGACTGAAAAGGTGCACAAAT  
AAGGTGTCTGTGAGAGAAATTAACCTTCCAGATACAGAGACACAACCTTC  
CCCAAGAGGTCCTCATTGCTCTGCCTTTTTTCTTTTTTTTGCTTGTTCT  
ACCATTAATAACAGAACTGATTATGACCTCAAAAGAGAGGAGAAAGCGA  
CTCTCCCCACCCTAGAGCTAGTTAACCACCATATCTTCTTAGATATCCTT  
GAGAGCAATGTAACCC

>Contig29

GTGAACTCGTTTTACCTGTGTAGCAGACCAAGCCGAGACAAAATCCNTC  
AGACACCAAATTAAGAAGGAAGGGCTTTATTGGGCCTGGAGCTGCGGCA  
AGACTCACGTCTCCAACAACCGAGCTCCCCGAGTGTGCAATTCCTGTCCC  
TTTTAAGGGCTCACAACCTTAAGGCGGTCCACATGAGAGAGTTCGTGATAG  
ATTGAGCAAGCAGGGGGTATGTGACTGGGGGCTGCATGCACCTGTAGTTA  
GAATGGAACAGAACATGACAGGGATCTTCACAGTGCTTTTCTTATGCAAA  
TAACCGATTAGATCAGGGGTCGATCTTTACCAGGGCCAGGGTGTGTCACC  
GGGCTGTCTGCTTGTGGATTTCATTTCTGCCTTTTAGTTATTACTTCTTT  
CTTTGGAGGCAGAAATTGGGCATAAGACAATATGAGGGGTGGTCTCCTCT  
CTTACCTGCGGGGAGTGAGCTCAAACCTCTTAAAGGAGTTACCTGCCTTC  
CATCATCAGGGAAGCAGGAAATCTTGCTTCTTGTGGAAGCAAGTAAA  
ACTCAAAACAAACAAAGAAAAAACAGGGAGTTGTACAGCAAAATAAACT  
TTTGATTTTGAACAAATTTGGGAGATCAGGAATTCCTCTGAAGGAGATGC  
TTTCAGACCTCAGCAAAATGTCTCTGTTGGTTTGAGCCATAAAGTTAGCTC  
ATGCTGGTACCAAAACACCAAGTAGGAGATTTGTCAAAGGTAAGAGGCATCT  
CCACTCAGAATCCCTTCGTGGTTACCAACATGTGAACCTTGGAAATCTGA  
GACAGGTCTCAGTTAATTTAGAAAAGTTTATTTTGCCACGGTTGAGGACAC  
CCACCCATGACAGAGCATCAGGAGGTCTGACCACATGTGCTCAGGGTGG  
TCTGAGCACAGCTTGGTTTTACACATTTTAGGGAGACATGAGACATCAGT  
GAATATATGTAAGATGTACACTGGTTCCCTCCAGAAAGGCAGAACAACTT  
GAAGCAGGGAGGGAGCTTCCAGGTCACAGGTAGGTGAGAGACAAACAATT  
GCATTCTTCTGAGTGTCTGATTAGCCTTTCCAAAGGAGGCAATCAGATAT  
GCATTTATCACAGTGAGCAGAGGGGTGACTTTGAATAGAATGGGAGGCAG  
GTTTGCCCTAAGCAGTTCCCAGCTTGACTTTTCCCTTTAGCTTAGTGATT  
TGGAGGCCCAAGATTTATTTTCTTCTACATCACTGTGGGCAGCTGACT  
AGGAAAGCTTTGTAGGACTGGTGGGCAGTGTGAGAGCCAGTGGGGGGTG  
GTGGTCTCTGTGCCAATGGTAGCAACCACCTGTGAGGCTGAGTAAACTCAT  
TTCCCAACCTCCTCTAGCAGCCCCAGTGGAGATACAGAGGAAGCAGACTA  
GCGATACAACCCAGCCTGAAGTTTTGTCTGGTGTGAGTGTAAATGGAATAAAA  
ATGGGAAGGGTGCTGAAGAGACCAGCAAGAAAATGGTTGAAGAGATGGGG  
CACAGAAATTAAGCTGGATCAAAAAGGACGGAAAAGCAGAAAGGGCCGAT  
AGAGAGAGGGGATATCTATGGGTTCCGATTCGAAAAGGACAAATCACT  
GGTGCTTTGAGAAGAGAGAGGGGTGAGAAAGCAGGAAGGCTGGAGGCTGTC  
ATCCAAGAGCGGACATCTGTGAACATGATTCCAAGAGTCACCAGACCAT  
GGGGGTGGCCAAAGGGAGTGCCTCTTCTCACTCTTAAATTCCTT  
GTACTCAAGATAATAAGTTCCCAGAAAGAGAAGTACCCATATTTAATTCAT  
CTGTGTCTTCTAGCAGTACTAAAAATATTATATGAAAGGTATCAAACCT  
TTGAGAATGTGTGCTGCTAAATTGTTAAGGATGCTGGAAAACCTCAAGACG  
TCCCTGATCCTGAGCCTGAGTATGAGCCTGTGGTGAGCCCAATGCAGGTC  
TCCATTACAGACAAAGGCCTCAGGGAACGGATGAGACCTAGGGACAGAGAT  
GCATGCTGGAGCAGCATTCCCCATCCCTACTGCAGCTCAGGCCAGCTGAC  
TGCTTTATGAGTAAACGTTACCAGGGAACACTTTGCAGTCTTAACACACA  
TGCCACCTGTGACCACTGATCCCTGTTGGGTGACCACTGACATCAGAGA  
TTCCGATGGCAGCAATGAAGACAAGGCTATCCTCATTAGGAAGGAAAGGAA  
GGAGGAGGGAGGAGGGCAAACGAATCTTTCCTGCTTGTCAACCACGTCCA  
TCTCTGTTAGGTGATTTCCCATGTGTGACTTTGTTTATCTTTATAATAAC  
TCTGAGAGGTAGGTCTTGATGTCCACATTTTGAACATGAGGACATCCAGC  
CAGGAAGTTGAGTTCTGGGGACATAGCTGAGAGGGCAAAGCTACATATAA  
ACCCCTCTTTGTTTTTCTGGCTTATCCACTGAGTGCCCCCTGCAATCCA  
CCAGCCCATTGTGAAGTGCATACTATAGGTAAGTTGGCACAGGAGGAGT

FIG. 3 (20 of 52)

22/118

GGATGTGGGCGATTTTG. JACAGCTCTCCAGGAACTTACACACTGGTGAG  
GAGGGCCAGGTATGTTCTGACCAGTCACAATCAAAGCAACCTCCTACTA  
ATCAGGGAGGCTTGGTACCTGGGGAATGCTATGTTGAAAGGTTCTTTTCT  
GGGTTTTAAATGATGGGTCTATTTCTTATTCTTAAGATTGCTTTTTTT  
CTGGCTAGAACTTAAAAGAAATTTTCAGTAAATTTCCCTTCCCTGGCAC  
AAAGTGAGCTTGAAATGAATTTCCAGGTGGCCTTGATACTTTAAATATT  
GCCTCCTATAAAATCAACCTTTAGAAGAAGGAAGTCAAAGAACATGCTAG  
ATTTACAAAGGTTAATTCCTTGAAATCCAGTTATCTACAGGACAATGTT  
GTCAAAGAAAAAATTATTTGGCCAGGCACGGCGGCTCATGCCTATAATCC  
CAGCACTTTGGGAGGCTGAGGCAGGTGATCACCTGAGGTCAGGAGTTCTGA  
GACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA  
AAAATTAGCCAGGTGTGGTGGTGGGCACCTGTAATCCAGCTACACGGGA  
GGCTGAGGCAGGAGAATCGCTTGAAACCGGGAGGAGGAAGTTGCAGTGAG  
CCAAGTTCAGCCACTGCACCCAGCCTGGGCAACAGAGCAAGACTTTGT  
CTCCAAAAAATAAATCAATGATATTTTAAATTCATGGTAAGGAA  
GATTTGATTTCAGAACCCAGACAGAGATATAGGAAACACTGCAATGGGAC  
TTTGCGGTGGGGGAGAGAGATTGAACACAACACTACATATACAGCACGGGCA  
AGGACATATTCATAGCCAGGAAGCAGAGCAAAGATCAGTGGATGCGAAAT  
TACTAAGAGGAAACATGAAAAATAAGGGAGCTTCTGCCTAAACCCACCTA  
ACCGGATCCTTGCTGAAGACAGGACAGGGTGATTGGACACCACTTTGGGG  
ATGGTGGAGGATGGGGAATCCAGTGAGATTTCAAGGGTGATGCGATATTG  
AACATACAAAGTTCTTGCTAAAAAAGGATTTTACAAGAAAGTGTAACAAT  
GTGCCTGGGACAAGGTGCAGGAGCCCGACGGAGATGTGGTCCAGCAGAGA  
ATATGTGCCGAGATGATAGGTGAGTTCTCTGACGAAGGATATATGCTGAT  
CCAGCCAGGGTGAAATGCTCAGAGAAAGCACGGAGGGGCTATGTCCGTTG  
CCCCAGTCTCCACGCGGTCAAATCTGATCCCCGTGTGAGTGTGGCCGTTT  
GTAGAAAGCAATCAGGGGGGGTCCCTCCCC

>Contig30

AATATATATTTTTTATANNATNTGAGACAGGTTCTCACTAGGTTGCCCAG  
GCTGGTCTTGAATTCCTGCCTTCAAGTGACTCTCCACCTTAGCCTACTG  
CATAGCTGGGATTACAGGCACAAACCACTGCATGCAGCTAACTTTGCTTC  
TCATTCAGCACTTTTTATTCCACTGATTATATGTATATGTATATCTGCA  
TCATCTCTCTCTCTCTCTCTCTCTCTCTCTATATATATATATATAT  
ATGGAAATATCTCTCTCTCTCTCTATATATATATATGGAATATATATCT  
CAGTCTCTCCTATCCTCCTTTAATCAGTTTTGCTATCCTGTCAATTCCTC  
CAACGAGTGTGATGTTGTGAAATATATATTTGTTCTTCATCTCCTGTTT  
CTGACATACAGCTTTTAAAAACCCCTTGAATCTCTGGAATAATAAGAGTG  
TCTTTTGCATGCTAATAGATGACTGCTGGCTGGCAGCCCCAATGCAGTAG  
CTTCATGATGGGGTTTGTACAGGAAAGACCAAGGCAGGATTGGAGACTT  
GAGACTGTTAGCCCCACTCCCCAACCCTGGAGGGAGTGGAGGGGCTGAA  
GGTGTGTGTCAGTCACCAATGGCCAATGGTTCCGTCATCATGTGTATGTA  
ATAAGCCACTCTTAAAAACCCAAAAGGACAGGGTTTGAAGGGCTCCC  
AGATAGCTGGACACATGAAGGTTCTTGAGGGTGGTGCCCCAGAGGGGCA  
TGAAGCTCCACACCCCTTCTCACATGCTTTGCTCTGCGCATCTCTCAT  
CTGGTGTTCATCTGTATCCTTTGTAATATCTTTAGAATAAACTGGTAAA  
CTTAAGTGTCTTCTGAGTTCTGTGAGCTGCTCTAGCAAATTCACGGAAC  
CCGAGGGAAGCAAACCCAGATTTATAGCCATCAGTCAGAAGCATAGGTGA  
CAACCTACCACTTGTAAGTGGCACCTGAAGTGGGAGGCAGTCTTGTGAGA  
CTGAGCCCTCAACCTGTGGGATCTAACGCTAACTCCAGGTAGATAGTGT  
GGAGTGAATTAGGACACCCAACTGGTGTCCGCTGCTGGAGGACTAGTGGT  
GGGAGAAATCCCCAAGCATTTCGGTGACTAGAGGTACAGAAGAACTCAG  
TGTTGAGGTGTGTGACAGTATGGTAGGGAAACTGCGTCTGGTTTTTCT  
CTTTACAATCAGTTAAATATTTAACACAAGTCTACTGTATATTAGTAAA  
AGGGTTACATTTTTTTCGTGTCCGTTTGAACCAAATCACTTGGGATACC  
TCAATCACTTTTTTTCGTGTCCGTTTGAACCAAATCACTTGGGATACC  
ATGAACCAGGCTGCAGCGTATTTCCAGGCCTTGAAAGCTTGAGGCCAT  
TTTGCCAGCCNTAATCCCTGTGAATACCAGGCTTCGTGGATTTAAAAAAT  
AGACTTGAGGCCAGGCCTGGTGGCTCACACCTGTAAGCCCAGCACTTTGG  
GAGGCAGAGGCGGATAGATCAAGGTTAGGAGTTGAGACCAGCGTGGC  
CAACATGGTGAAACCCCGTCTCTACTAAATATACAAAAAATAAGCCG

FIG. 3 (22 of 52)

CAACACCTCTCACTAAAGAGAAAGAAATAAAAAAGAAAATTAAAATCTGC  
CGCAATGCCACACAGTCATTGAATAACTGCATGTGTACAGCACTTGGTT  
ACTTTTACATACTTCATATTTTAGCCTTCATAGCAGCTCACAGGGGTGGA  
TTTAATTTTGTAGTCCAACCTCCTGTCACGGTGCCTGGCACAAGTATAATAA  
ATGTTCTGTGAATAAATGACCCCTCTTTTATAGATGAGGAAATCGAGGCTCA  
AGGAGAACAAGCAATGTAATGTCCCCCTCCTGTTGAGCCATCTGCCTTTC  
ACGCCACTGAATGCAGTAGTCCTCAGTGCCCTGAACTTGACCCCTCTTCTG  
CTTTTCGGACTGGTCCTTCTAATCCCGTTGTGACTCACTACACCACCTCT  
CCTGCATATGACATCTACATTTTAAAACAAACCGTATGGAAATAACACAT  
TAGTCGGCTTGTTCCTCCACCCCGCAAAAAAAGGCCTCTTTATAACA  
GAAACTTCTCAGGCTGGTAGGGGAATTTTATTCCCCCATTTATGGTAGAA  
AGGCCCTAACCTTGGACCTCACGCCATAGCTATTACATGGGGGAATGAT  
GAATAACATGGGGAGCAGCATGTAAATATCATTGAGCCGTAGTCCAGACC  
TATAACACATC

>Contig31

GGGGGAGCTGCATGTGCCTGTGCGAGATCTGGGGGAGGAACAGGAAGATCA  
AGAGTTCTGTGTAGGACATGTTAAGTTGAAGGTGCTTACAGGATAGCCAG  
ATGAAGCATCAGGTGTGCAGTCAAAGATATGAGTCTGGAGCAGCACATCC  
TAAGTCACTCCTGTCACCAACACAGAAGTTCAGGCCACTCACTTGAGCT  
CTCCCAAATAGTTTCCAAGTGTGCTATTATGTTAATAACCTATGAGCTTGAA  
CACCAGATTCAAACCCCACTGCATGGCTTTTAAAGACCATCTCAAGGGCT  
TGACACTCCAGGGAGCCAACTAAAGATGCCTGGTCTTACCATCAACCTCC  
ACCCCATTTTTTATAGAAAATGTTTCTACCTGTCTAAGGCAGGGTCTTG  
CCCCACTCCAGGCCCTTTAGATCCCCAATATTCTCCTCCCTGAACCA  
AAACCCCTCATCATCTTCCAGCATGGGTGGGGCCTCCATTCTTGCTTCTGC  
TCCCTTGAGCAGAAGCAAGTTTCTCCCACTTGACCTGATTCTCCTCCTA  
AGTACCAGTCACTGCTTTGTTTCTGGAATGAGAGAAAAAGACAGAGTGAG  
AGAGACAATCCGAACCTCTGCTCACTCACAGCTAGGCTGGGCATCTGGG  
AGGATGGCTGTGTCATGGGAACCTGGGAAAGCCACACCCTTGGCACCC  
TGGTCACCCACCTGTCTCCCTGGCAGATTCCGCACTGCTCTCTTGACCC  
TCTACCAGGGCTAACCGGCCTGCTCACTCTCCCCAGCATGTCTTCCCAG  
CCCCTCTCTAATTATTACATTCCCTTCACATAAACTGCCCTTCTCTCCC  
AATCACCACATGTTCACTTCCCACCCAGCTGTCAAAGTCTGGCTCAACCT  
CATTCTTGAAAAGGAAAAAACAACAACAACAACAACAAGCAAAAA  
ACCTATGATGGATTAAGAACACACTTCATTCCAGGAACATGCTTATCTCC  
TCTAACTCTCACAACAACCTACAGCAGGTAGGTGTTATCACACCCATCTCT  
CAGGTGAGAAAAACAGGCTCAACGAGTGCAGGAGGACACAGCAAGTCAGTG  
ACAAAGCTTAAATTCAAGCCCAAGCCTGTTGGCAACCAACGTCTGTACCC  
TTGATAGCTACCTCATTTACCACCAATCCAGTGGCCTCAGGCCTGGCTG  
CACACTGGGATCACCTGGTGCCAGACCACATCTTAGACCAGTCATACAG  
AATCTCTTGGGCTGGGATCCTCCACGGTACATTTTAAGGGTCCCCAGGTG  
AGTTCCACCATGGACCCAGAATTGAGGACCCAATACCGTATACCATCTCC  
TTCTTCATCTCTTCTAAGGCATCTCTTACTCGCTGTGCACTCCCATACCA  
CTTTGTTCAATCATCCAATCATTCACTTATTGAGTCAGTTAGTCAGGAGC  
TACTCACTAGTCCCCTGCCAGGTCTTAGTCATGACATAGGGCTCTGGGGA  
CCAACAAGAAGCAGGACCCATGCCCTCCTGCTCTCATGGAGCTTGCTCTGC  
AGCAGAGGAAGCAGTCAGTGAGATGTAGCAAATGTGAAATGTGCACAGAT  
GGGAAAAGCAAACTTTAAAACCTTTTAGGACAAAATACACAAGAAATCTT  
TGCAACTTTGGGACAGGAAGGAACAACATTCTTACACATGACACCAAAG  
GAATCAACCATAAATAAAAAAGGTGATCAATTTGACCTCATTTAAGTGTTA  
AGCTTTTTTTCATTGAGAGACACCATTAAAAATTAAAAATACATGCCACAA  
ACTGGGATACAATATTTACAACACTTATGTCTCACAAGGATTAGTTTTTC  
AGAATATATAAAGAACTCCCGGCCGGGTATGGCCGCGCACGCTGGAATCT  
CAGCACTTTGGGAGGCCAGCGGATCACATGAGGTGAGGAGTTCAAGACCA  
GCCTGGCCAACATGGCAAACTCCGTCTCTACTAAAAATACAAAAATTAG  
CCAGGCATGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAACTGAGG  
CAGGGAATCACTTGAGCCAGAAAACAGAAGTTGCAGTGAGCTGAGCTC  
ACATCACTGTAAGCCTCGGTGACAGAGTAAGACTGTCAAAAAAACGAAAA  
CAAAAAACAAAACTCCTACAAATAAATAAGAAAAAATAGCCAGCAGGA  
AAAAGTATATACATTTCTATAAAGAATAAATAACATTCTGTGAGTTTCTA

ACATATATTTTTTAAGAGTAAATACAAATGGTTAGGAAACATTTTTTAAA  
ATGCCCAACCTCATTAAAAATTATAGAAGTGAAAATTAAGCCACAATAAG  
ATACGATTTTATACCAAATACAGTGTCAACACTTTGCAAGTCTGACCTCA  
CCAAGTGTACCAGACGTGTGCACTGACGTGGCTGCTGAGATACTGATGG  
TGGGTCTGTAAATCTGTACTACAAACAATTGCAATAAAATGTAATAAATA  
TACAATAGGTGGAGCAGGAAGTGACCTGCAACCATATAGCAGATAGGGCA  
GGAAAAAGCCTATGAAAGCTGACATCAAAGGGATAAGTTCCAGTTACCCA  
GCTGAAGGGAAGGAGGGTGTTCAGATAGAGGAAGGATAAGCATGACCTA  
TTCAAGGCCAGTGAAAGAAGCGTGCAACGGCCAAGTCAGGAGAACCTGAA  
ATTGTGTCAAAGAGCTTGGATGCAAAGAGCCGTGGGAGACTATTGGGGGT  
TTTAAGCAGGGATATAATATTCAATTCAAGCATGCAGTAAAAGGTCAGTGG  
CACCTGCCATGGGCCAGGACTCGGGCTCTACATGATTGCGTCTGTTTTGG  
AAATATCACCTGGCTGTGAGATGAAGAACAGGTAGGAGGGTCACAAAAC  
TTGAAGCAGAGAGACTGTTGAGGAAGTAAGCTGTTTTTGTGTGGACTGTG  
GCAATCACAGAGGCAGAGGATATAAATGCACAGAGACACAAGGCATGTGG  
GAGGCAGAAGGAATCAAATACAATGAGTGATCAGATGTGGGGTTAGAGTG  
GTGAGTGAGAAGACATACTCAAGGTGACACGCCAGGTATCTGGGTGGAT  
GGTAAGACATTCATGGACTAGGATCGAGGAANGAGGTGGGGAATGGGACC  
ATACCTGCAGTTTATAAGGGGTGGACGAGGGAAGATTATGCGGGAGACTG  
AGAGAGGAATAGACAAAGGAATCCCGGTGCAGTATTACAGAACTGGGGT  
GGGAGGGGTGTANTTCAAAAAGGAAAGAAAATTGTCAAATAGTATGAA  
ATGCTGCAGAGAAAACCTCACGGATTTTTTTTTTAAGCTTAGAATTATTCAT  
TGACTATGTGAATAAGAATAACTTTTTATGAAAGAAGTTTTGCTTAAGTAG  
TAGGAAGAAGCAAAATTGTTGAGGGCTGATGAGTGGGAGGAGAAGTAATT  
GAAGGCACCTCTTCAAGAGAAACAAAGCAGAAGGTGAGGAGAATACTAAT  
GAAGGAGTTACGGCCTTCACTATTTTTGTTTTGCTTTAGATAAGCAAGACT  
TGAGTGGGTCTGGTGAGGAGAAACAAGTAGAGTACAAAGTTAAAGGAGAG  
ACAGACAGAGATAGAGATAGGGACAGAGAGAGAGACAGAGACAGAGCACA  
AAAGAGCAAGGTCCCTGAGAACACGGGCCTTCTGTTTAAACCCAGCCAG  
ATGTATTGCAATTCAATTCCAGTACTAACCACCCAGAGTTTGTGTAGACT  
CTACAAGTTAAAGAGCATGGTCCCCAACAAGACTGCTTCTACGTCAGATG  
CCAGGCACACTTCAGGGGTCCCCAAGCCACTCATGTTTTTTGAATGACTG  
CCATAAGTTCAAAAATTCCCACAATTCTCTCAGATTCAATAACTGGGTAT  
AACCCTCATAGAACTCAAGAAAATGCTATCATTATTATTACAATTTTAT  
TATAAAGGATACAAATCAGAAGGACTAGCCAAATGAGGAGACACATAGAG  
AGAGGACTAGTAAAAAACAGAGCTTCTGCGTCTACCTTCAAGGAATCAG  
GATGCACCACCTCCAGCACATCAAGTGCTCATCAACCAGGAAGTTCCT  
CTGAGCTCCAATGTCCAGAGATTTTAGGGAGGATTCATTACATAGGTATC  
ATTGATTAAATCATTTGGCCATGTACTTGAACCTCAATCTCCAGTGTCCCTC  
TTCTCCCTAGAGGTCTGAAGGGTTGGCTAATATCATGTGGCTCAAAGCCC  
CAACTCTAATTACCTTTTTTGGTCTTTTCAGGGACTAGACCCATCCTGAA  
GCTATCTACAGGCCCTGCCATGAGTTAGCTCATTAACATAACAAAGACAC  
TTATATTACTCAGAAAATTCCAACAGTTTTTAGAAGCTCCATGTCAGGAAC  
CTGGGACATAGATCAAATTCTTTTTTTTTTTTTTTTTTTGGAGACAGGGT  
CTTGCTGTGTTGCCAGGCTAGAGTGCAAGGACAGATCACAGCTCAATGC  
AGCTTCAACTTCCAGGCTTAAGTGACCTTTCCACCTTAACCTTCCAAGT  
ATCTGGGACCACAGAAAATGGCTAATTATCCTGGCTGATTTTTAACTTT  
TTTTTTTTTAGGGATGGGATCGCCCTGTGTTGCCAAGGTGGTCTCAA  
CTCCTGGGTTCAAGCAATCATTCTGCCCTGGCCTCTGTGATGGTTAATAC  
TGAGTGTCAACTTGATTGGATTGAAGGATACAAAGTATTATTTTTGGGTG  
TGTCTGTGAGGGTGTGGCAAAGGAGATTACATTTGAGTCAGTGGACTGG  
GAAAGTCCACCCTTTCCAGTGGACTGGGAGACCCACCCTCAATCCAGGT  
AAACACAATCTAATCAGCTGCCAGTGTGGTCAGAATAAAAGGAGGCAGAA  
GAACAGGGAAACACTAGACTGGCTTAGTCTTCCAGCCTACATCTTTCTCT  
CATGCTGAATGCTTCCCTACCTCGAACATCAGCCTCCAAGTTCTTCAGTT  
TTTGGACTCTTGGACCTTCAACCACAGATTGAAGACTGCAGTGTGGCTT  
CCCTGATTTTTGAGGTTTTGGGACTCAGACTGGCTTCTTGCTCCTCAGCT  
TGCAGATGGCCAATTGTGGGACTTTAAGTTGTGATCATGTGAGTCAATAT  
TCCTTAATAAAGTCAATATATATATATGTATCAGACATATATATATATC  
CTATTGTATATTATATACAGATATATAATATCCTATTATATACAGATATA

TAATATCCTATTATATACAGGTATATATATATATATGTATCATATATA  
TATCCTATTGGTTCTATCCCTCTTGAGAATCCTGACTAATACAGCCTCCC  
AAAATGCTGAGATTACAGGAGTGAGCCACAGCCACCATGCCCAGCCCCAA  
ATTCTTAATTATACAACAATGGGTCCAGAGATCAGGGCCTGGGTAGGATG  
CAGCAATAAGAAAACAGATGGTGGATGGGGACACATGTTGGAAGTGTGGC  
AGGACATGGCTGAGGGAACTCATAGGATGGTGTCTATTTTCATGGCTGAG  
TGTGAGGAACAGCATAAGGTCAAAATTTTCAGGTCAATGGTGAGTTTTTTA  
AATTGTTGCTGTGAACCCCAAAAATCTGACCCAGGTCTCAGTTAATTTAG  
AAAGTCTATTTTTCCAAGGTTGAGAACACCCACCCACTCAGACAAGAGC  
ATCAGGAGGTCTTGACCACATGTGCCCAAGGTGGTAAGAGCACAGCTTGG  
TTTTATATATTTTAGGGAGACGTAAGTCATCAATCAATATATGTAAGATG  
TACACTGGTCTGCCTAGAAAGGCAGGACAACCTGAAGCAGGGAGGGGGC  
TTCCATGTACAGGTAGGTGAGAGACAAACAGTTGCATTCTTTGAGTTTC  
TGATTATCCTTTCCAAGGAGGCAATCAGATGTGCAATTATCTCAGTGAG  
CAGAGGGATGACTTTGAATAGAAAGACAGGCAGGTTTGCCCTAAGAAGTT  
CCCAGCTTGACTTTTTCTTTAGCTTTGTGATTTGGAGGGCCCAAGATTT  
ATTTTCCTTTACATTTCCCCCCTTTCTTTTTAAGAATCTTTTAAAGAA  
AGCTTTTAAAAAGAAAATGAGTCTCTGGTCCCAGGTTTCATCTGAATTCT  
CGAGGGGAGGATGGTTTATCCTAAACGGGTGGTCTGAATTTTGAGAAAG  
TGCATTGTAC

>Contig32

AAAAGCCATACGAATGAGGAAGAATTAAGGGCCAGAACAAAACAAGAAGA  
TGAGGGAAAGTTTGGAACTTCTTAGAGACTGGCTAAATGGTTGTGACCAA  
AATGCTGATAGTGATACGGACAATGAAGTCCAGGGTGACAAAGTCTCAGA  
TGGAATGGGGAAATTTGTTGGGAACTGGGCAAAGGTCAACCTTGCTATGA  
CTCAGCAAAGAAATGGGTGCATTGTGTTTCATGTCTGGGGATCTGTGGA  
AGTTTGAATGTAAGAGTGATGACTTACGGTAGGGTATCTAGTGGAAGAAA  
CCTCTAAGCAACAAAGTGTGTTGCTTAGAAATTTCTTTCTTTCTTTTTTT  
TTTTTTTTTGAGCTGGAGTTTGTCTGTGTCGCCCAGGCTGGAGCGCAGTG  
GCGCAATCTTGGCTCACTTCAAGCTCTGTCTCTGGGTTTCATGCCATTCT  
CCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCTGCCACCATAC  
CTGGCTAATTTTTTAGTATTTTAGTAGAGACGAGGTTTCACCATGTTAGC  
CAAGATGGTCTCAATCTTCTGACCTCGTGATCCACCCGCTTGCCCTCCC  
AAAATGCTGGGGTTACAAGCATGAGCCACCCCGCCTGGCCTGCTTAGAAA  
TTTCTAAGCCAGGATATGGCCTGTCTGCTTCTAACAGCCTGTGCTCAGGG  
GTAAAGAAATGACTTAAAGTTGGAACCTATGTTTAAATGGAAGTAGAGT  
CTAAAAATTTGAAAATTTGCAGCCTGGCCTTGTTGGCAGAGAAAGAATCC  
AAGTAGGCTGCAGAGCAATCATTGCTAGAGAGATTAGCATGACTAAAAGG  
GAGCCAAGTGCTAATATTCAAGACAATGTTAAAAAGGCCCTGAGGGCATT  
TCAGAGATCTATGAAGCAGCCCCCTCCCATCACAGGTGCAGAGGTTTGGTG  
CACTAGGCCCAGAGGTTTTATGGGCCANNGCCAGGGCCCACTGCTATGC  
ACAGCTTTGGGACACTGCTGCCCGCATCCAGGCCACTCTGCTCTGGCTCC  
ACCTTTGGCTCAAACGGGCCAAGATAGAGCTTGGACCACTGCTCCCGAGG  
GCACAAGCCATAAGCCTTGGTGGTTTCCATGTGGTGTAAAGCCTGCAGGT  
GCCCAGAATGCAAGATTGAGGGAGCTTGGGGCACTTCCACCTAAATTTAG  
AGGATGTGTGAGAAACCTAGGTTCCCAGGCAGAAGCATGATACAGGGGC  
AGAGCCCTTGCAAGAGAACCTCTACTAGGGCAATGCCAAAGGAAAATGTGG  
GGTTGGAGTCTCACACATGGTCCCCACTGGGGCACTACCTGGTGATACT  
GTGGGAATGGGGCTGCTGCCCTCCAGACCCCAAGATGGTAGATGCACTGG  
CAGCTGGCACCCCTGAGCCTGGAAAAGCTGCAGGCACTCAACTCCAACCCA  
TGAGATCAGCCACATGGGCTACTCCAGGGAAGCCACAGAGGCAGGGCT  
GTCTAAGGCCCTTGGGAGCCTACCCCTTGAACCAGCTTGCAAGACATGGAA  
TCAAAGATTATGTTGCAGCTTTAAGGCTTAATGTTTTCCCTGTCAATTC  
AGGCTTGTGTGGGACCTGTTGCTTTTTTTTTTTTTTTTTTTTTTTGGT  
CACAGGTGTTTGAACCAGAACATCCATCTGAATAGGGGCTGGGTAAA  
ATAAGGCTGAGACCTACTGAGCTGCATTCTAGGAGGTAGGAATTCTAA  
ATCAGAGGAGGATAGGAGGTGGGCACAAGATACAGGTAGCGAAGACCT  
CGCTGATAAAATAAGTTGCAGTAAAGAAGCCAGCCAAAACCTCACAAAGCC  
AAAATGGTGATATGGTTTGGCTCTATGTCCCCACCCAAATCTCATCTCAA  
ATTATAATCCCATAATCCCCACATGTTGAGGGGAGGACCTGGTTGGAGG

TGATTGGATTATGGAGGCAATTTCCCCCATGCTGTTCTGGTGATACTGAG  
TGAGTTCTCATAAGATCTAATGGTTTTATAAGTGGTTGGAAGTTCCTCCT  
ACACACATGCTCACACTCTCTCCTGCAGCTTTATGAAGAAGGTACTTGCT  
TTCCTTTCTGCCATGATTGTAAGTTTCTGAGGCTTCCCAGCTATGCAGA  
ACTGTGAGTCAATTAAACCCGTTTTCTTTATACATTACCAGTCTTGGGCA  
GTTCTTTACAGCAGTGTGAGAACTGCTGGCGATGAGAGTGACCTCTGGTT  
GTCTCACTGCTCATTATATGCTAATTATAATGTATTAGCATGCCAAAAG  
ACACTCCCACCATGACCCCAACAGTCATGCCTGTGCCGGTCTCAGCACCA  
TGACAGTTTACAGATGGCATAGCAACGTCTAAAAGGTACCCCATATGGAC  
TAACAAGGGGAGGAACCCCTCAGCTCTGGGAAGTGCCCTACCTCGTTCCAG  
AAAGCTTGTGAATAATCCACTGCTTGTTTAACATATAATTAGAAATAAC  
TATTAAGCATCCTTAGTTTACAGAGCCCAAGCTGCTGTTCTGCCTATGGAG  
TAGCCATTCTTTATTCCGTTACTTTCTTAATAAAATTGCTTTTACTTTAC  
TGTATGTACTCGCCTGGAATTCTTTCTGTACGAGGTCCAGAGCCCTCTC  
TTGGGTCTGGATCGGGACCCCTTTCTGGTAACATTTTGACCAATTTCTCC  
CTTCTGGAATGGGAATGTTTACACAATGACTGTATCACTTTTGAATCTTG  
GAAGTAAATAATTTGTTTTTGACTTTACAGCCTCATAGGTGGAAGGAACT  
TGACTTGAATTTAGATGAGACTTTGGACTTTGGGACTTTTGGGTTGGGG  
CTGGAATGAGTTAAAAGTTGGGGGATTATTGGGAAGGCACGATTTTATT  
TTGCAATATGAGAAGCACATGAGATTGGGGGACCAAGGGTGGAAATAATA  
TGGTTTGGATGTTTGCCCCCTCCAAATCTCACATTGAAATGTAATCCCCA  
GTGTTGAAGTGAGGCCTGCTGGAAATGTTTGGATTACAAGGCTGTGAG  
CACATTGGATAAGACGTGTAGGNCCC

>Contig33

CGCAGCTCGCTGGTTAATTCTGTGGCTCCTGTGACCACTATTATAGCACC  
AGGTCTATGACCAGGAGAATTAGACTGGCATTAAATCAGAATAAGAGATT  
TTGCACCTGCAATAGACCTTATGACACCTAACCAACCCCATTTTACAA  
TTAAACAGGAACAGAGGGAATACTTTATCCAACCTCACACAAGCTGCTTTC  
CTCCAGATCCATGCTTTTTTTCGCTTTATTATTTTTTAGAGATGGGGGCT  
TCACTATGTTGCCACACTGGACTAAAACCTCTGGGCCTCAAGTGATTGTC  
CTGCCTCAGCCTCCTGAATAGCTGGGACTACAGGGGCATGCCATCACACC  
TAGTTCATTTCCTCTATTTAAAATATACATGGCTTAAACCTCCAACCTGGGA  
ACCCAAAACATTCAATTTGCTAAGAGTCTGGTGTCTACCACCTGAACTAG  
GCTGGCCACAGGAATTATAAAAGCTGAGAAATCTTTAATAATAGTAACC  
AGGCAACACCATTGAAGGCTCATATGTAAAAATCCATGCCTTCCTTTCTC  
CCAATCTCCATTCCCAAACCTAGCCACTGGCTTCTGGCTGAGGCCTTACG  
CATACCTCCCGGGGCTTGACACACCTTCTTCTACAGAAGACACACCTTG  
GGCATATCTTACAGAAGACGAGCTTCTCTCTGGTCCTTGGTAGAGGGCT  
ACTTTACTGTAACAGGGCCAGGGTGGAGAATTCTCTCCTGAAGCTCCATC  
CCCTCTATAGGAATGTGTTGACAATATTCAGAAGAGTAGGAGGATCAAG  
ACTTCTTTGTGCTCAAATACCACTGTTCTCTTCTTACCCTGCCCTAACC  
AGGAGCTTGTACCCCCAACTCTGAGGTGATTTATGCCTTAATCAAGCAA  
ACTTCCCTCTTCAGAAAAGATGGCTCATTTTCCCTCAAAGTTGCCAGGA  
GCTGCCAAGTATTCTGCCAATTCACCCTGGAGCACAAATCAACAAATTCAG  
CCAGAACACAACTACAGCTACTATTAGAACTATTATTATTAATAAATTC  
TCTCCAAATCTAGCCCCCTGACTTCGGATTTCACGATTTCTCCCTTCCTC  
CTAGAAACTTGATAAGTTTCCCGCGCTTCCCTTTTTCTAAGACTACATGT  
TTGTCTCTTATAAAGCAAAGGGGTGAATAAATGAACCAAATCAATAACT  
TCTGGAATATCTGCAAACAATAATATCAGCTATGCCATCTTTCACCTA  
TTTTAGCCAGTATCGAGTTGAATGAACATAGAAAAATACAAAACCTGAATT  
CTTCCCTGTAAATTCCTCGTTTTGACGACGCACTTGTAGCCACGTAGCCA  
CGCCTACTTAAGACAATTACAAAAGGCGAAGAAGACTGACTCAGGCTTAA  
GCTGCCAGCCAGAGAGGGAGTCATTTTATTGGCGTTTGAGTCAGCAAAGG  
TATTGTCCTCACATCTCTGGCTATTAAAGTATTTCTGTTGTTGTTTTCT  
TCTTTGGCTGTTTTCTCTCACATTGCCTTCTCTAAAGCTACAGCCTCTCC  
TTTCTTTTCTGTTCCCTCCCTGGTTTGGTATGTGACCTAGAATTACAGTC  
AGATTTTCAGAAAATGATTCTCTCATTTTGTGCTGATAAGGACTGATTCTTT  
TACTGAGGGACGGCAGAACTAGTTTCTATGAGGGCATGGGTGAATACAA  
CTGAGGCTTCTCATGGGAGGGAATCTCTACTATCCAAAATTATTAGGAGA  
AAATTGAAAATTTCCAACCTCTGTCTCTCTTACCTCTGTGTAAGGCAAA

FIG. 3 (26 of 52)

28/118

TACCTTATTCTTGTGGTGTTTTGTAACCTCTTCAAACCTTTCATTGATTG  
AATGCCTGTTCTGGCAATACATTAGGTTGGGCACATAAGGAATACCAACA  
TAAATAAAACATTCTAAAAGAAGTTTACGATCTAATAAAGGAGACAGGTA  
CATAGCAAACCTAATTCAAAGGAGCTAGAAGATGGAGAAAATGCTGAATGT  
GGACTAAGTCATTCAACAAAGTTTTTCAGGAAGCACAAAGAGGAGGGGCTC  
CCCTCACAGATATCTGGATTAGAGGCTGGCTGAGCTGATGGTGGCTGGTG  
TCTCTGTTGCAAAAGTCAAGATGGCCAAAGTTCAGACATGTTTGAAGA  
CCTGAAGAATCTGTTACAGGTAAGGAATAAGATTTATCTCTTGTGATTTAA  
TGAGGGTTTTCAAGGCTCACAAAATCCAGCTAGGCATAACAGTGGCCAGC  
ATGGGGGCGAGGCCGAGAGGTTGTAAAGATGTGTACTAGTCCTGAAGTC  
AGAGCAGGTTCAAGAGAAGACCCAGAAAACTAAGCATTCAAGCATGTTAAA  
CTGAGATTACATTGGCAGGGAGACCGCCATTTTAGAAAAATTATTTTTGA  
GGTCTGCTGAGCCCTACATGAATATCAGCATCAACTTAGACACAGCCTCT  
GTTGAGATCACATGCCCTGATATAAGAATGGGTTTTACTGGTCCATTCTC  
AGGAAAACCTTGATCTCATTCAAGGAACAGGAAATGGCTCCACAGCAAGCTG  
GGCATGTGAACCTCACATATGCAGGCAAATCTCACTCAGATGTFAGAAGAAA  
GGTAAATGAACACAAAGATAAAAATTACGGAACATATTAACCTAACATGAT  
GTTTCCATTATCTGTAGTAAATACTAACACAACTAGGCTGTCAAATTT  
TGCCTGGATATTTTACTAAGTATAAATTATGAAATCTGTTTTAGTGAATA  
CATGAAAGTAATGTGTAAACATATAATCTATTTGGTTAAAAATAAAAGGAA  
GTGCTTCAAACCTTTCTTTTCTCTAAAGGAGCTTAACATTCTTCCCTGA  
ACTTCAATTAAAGCTCTTCAATTTGTTAGCCAAGTCCAATTTTTACAGAT  
AAAGCACAGGTAAAGCTCAAAGCCTGTCTTGATGACTACTAATTCAGAT  
TAGTAAGATATGAATTACTCTACCTATGTGTATGTGTAGAAGTCCTTAAA  
TTTCAAAGATGACAGTAATGGCCATGTGTATGTGTGTGACCCACAACAT  
CATGGTCATTAAAGTACATTGGCCAGAGACCACACTGAAATAACAACAAT  
TACATTCTCATCTCTTATTTTGACAGTGAAAATGAAGAAGACAGTTCCT  
CCATTGATCATCTGTCTCTGAATCAGGTAAGCAAATGACTGTAATTCTCA  
TGGGACTGCTATTCTTACACAGTGGTTTTCTTCATCCAAAGAGAACAGCAA  
TGACTTGAATCTTAAATACTTTTGTGTTTACCCTCACTAGAGGTCCAGAGA  
CCTGTCTTTTATTATAAGTGAGACCAGCTGCCTCTCTAACTAATAGTTG  
ATGTGCATTGGCTTCTCCAGAACAGAGCAGAACTATCCCAAATCCCTGA  
GAACTGGAGTCTCCTGGGGCAGGCTTCATCAGGATGTTAGTTATGCCATC  
CTGAGAAAAGGCCCGCAGGCCGCTTCACCAGGTGTCTGTCTCCTAATGTG  
ATGTGTTGTGGTTGTCTTCTCTGACACCAGCATCAGAGGTTAGAGAAAGT  
CTCCAAACATGAAGCTGAGAGAGAGGAAGCAAGCCAGTTGAAAGTGAGAA  
GTCTACAGCCACTCATCAATCTGTGTTATTGTGTTTGGAGACCACAAATA  
GACACTATAAGTACTGCCTAGTATGTCTTCAGTACTGGCTTTAAAGCTG  
TCCCCAAAGGAGTATTTCTAAAATATTTTGAAGCATTGTTAAGCAGATTTT  
TAACCTCCTGAGAGGGAATAATTGGAAAGCTACCACTCACTACAATCAT  
TGTTAACCTATTTAGTTACAACATCTCATTTTTGAAGCATGCAAATAAATG  
AAAAATCTTCTTAAAAAATCATCTTTTTATCCTGGAAGGAGGAAGGAAG  
GTGAGACAAAAGGGAGAGAGGGAGGGAAGCCTAATGAAACACCAGTTACC  
TAAGACCAGAAATGGAGATCTTCTCACTACCTCTGTTGAATACAGCACCT  
ACTGAAAGAACTTTTATTCCCTGACCATGAACAGCCTCTCAGCTTCTGTT  
TTCCTTCTCTACAGAAATCCTTCTATCATGTAAGNTATGGCCCACTCCAT  
GAAGGCTGCATGGATCAATCTGTGTCTCTGAGTATCTCTGAAACCTCTAA  
AACATCCAAGCTTACCTTCAAGGAGAGCATGGTGGTAGTAGCAACCAACG  
GGAAGGTTCTGAAGAAGAGACGGTTGAGTTTAAAGCCAATCCATCACTGAT  
GATGACCTGGAGGCCATCGCCAATGACTCAGAGGAAGGTAAGGGGTCAAG  
CACAATAATATCTTTCTTTTACAGTTTTAAGCAAGTAGGGACAGTAGAAT  
TTAGGGGAAAATTAACGTGGAGTCAGAATAACAAGAAGACAACCAAGCA  
TTAGTCTGGTAACTATACAGAGGAAAATTAATTTTTATCCTTCTCCAGGA  
GGGAGAAATGAGCAGTGGCCTGAATCGAGAATACTTGCTCACAGCCATTA  
TTTCTTAGCCATATTGTAAAGGTCGTGTGACTTTTAGCCTTTTCAGGAGAA  
AGCAGTAATAAGACCACTTACGAGCTATGTTCTCTCATACTAACTATGC  
CTCCTTGGTCACTGTACATAATCTTTTCGTGATTGATTTCTCTACTGT  
AAAATGGAGATAATCAGAATCCCCCACTCATTGGATTGTTGTAAAGATTA  
AGAGTCTCAGGCTTTACAGACTGAGCTAGCTGGGCCCTCCTGACTGTTAT  
AAAGATTAAATGAGTCAACATCCCTAACTTCTGACTAGAATAATGTCT

FIG. 3 (27 of 52)

29/1/98

GGTACAAAGTAAGCACC AATAAATGTTAGCTATTACTATCATTATTA  
ATTATTTTATTTTTTTTTTTTGGAGATGGAGTCTCACTCTGTTGCCAGGC  
TGGAGTGCAAGTGGCGCAATCTTGGCTCACTGCAAGCTCTGCCTCCTGGGT  
TCACGCCATTTTCTGCCTCAGCCTCCCGAGTAGCTGGGACAACAGGCAT  
GTGCCACCATGCCCAGCTAATTTTTTTGTATTTTGTAGAGATGGGGTT  
TCACTGTGTTAGCCAGGATGGTCTCTATTTCTGATCTCATGATCCGCCT  
GCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCG  
GCTTATTATTATTATTACTACTACTACTACCTATATGAATACTACCA  
GCAATACTAATTTATTAATGACTGGATTATGTCTAAACCTCACAAGAATC  
CTACCTTCTCATTTTACATAAAAGGAACTAAGCTCATTGAGATAGGTAA  
ACTGCCCAATGGCATAACATCTGTAAGTGGGAGAGCCTCAAATCTAATTCA  
GTTCTACCTGAGTAAAAAATCATGGTTTCTCCTCCATCCCTTTACTGTA  
CAAGCCTCCACATGAACATAAACCCTAATTTCTGTTTTTAAGATAATA  
CCTAAGCAATAACGCATGTTTACCTAGAAGGTTTTAAAATGTAACACAAT  
ATAAGAAAATAAAAATCACTCATATCGTCAGTGAGAGTTTACTACTGCCA  
GCACTATGGTATGTTTCTTAAATCTTTGCTATACACATACTACATGT  
GAACAAATATGTCTAACATCAAGACCACACTATTACAACCTTTATATCCA  
GCTTTTCTGACTTAGCAATGTATTGATGACATTATGCATGCTTAGACCTC  
C

>Contig34

GTATTTCTATTCTCGGTTATAACACAATCACAGTGATTTGTCTATATCTTTC  
CAGGATTTGTTAATTTCACTTCTTCAGCTGTTTCCCCCTTGTTGGCTGGA  
ACTGATTTTCTATCTTCTGGGAGAATCTTCAGCAAGCCAACTCAGGATTT  
GTTGGGTGCATTTTGTCAAGTCTAGGACCCAGGCTCTGGGTGACTGATTT  
CCTCTAATTACCGAGCAATGTAAATGAGGAAGTCTGATTGTGTAAAGGT  
GTTAAACTTTTGTGTGACGGCAAACTTTAATACCATGAATAGAGATTCC  
AGAATTTTCCAACCTTCTAACGGGATTCTTTCACTCCCTGACATTAGAAT  
GTTAGAAAATCTACCACAAAACATCTGTGAGGCTATCCTACAAGGCCCGT  
TTTTCAAATAGGTTTTTACAAGGATTGCTATTTGGGATGATAGTTTCAG  
AAAGGCGCTATCAAAGTTAATTGATGATGTGTGCAAGCTGAAAGTTATAT  
GTTAGAACTAGCAGTGATTTCAAATATCCCTTTTAGGCTTTTGTCTAA  
TATATCTGCTCATTTTCAAAGTTCCCAATATTATAAACTTTTTAAAGCA  
GAAAGAAGAACCTCCATTTCTGCTGGCCCCCTTCCCTGTTCAACTAAAAA  
GTATTTTCCCAGGCAATGCTATCCCAGGACTCACACTCCATCCATCCATC  
ACCTACCATAAGTTCTTTGAAGGGCTCATTCTGAGCGCTTCTGAGTGCC  
TGGGATCTGTTATTTCTCTCCATTTCTGCTGCTGCATGGTAGTCCAAGTC  
CTCCTCCCTTTTCCCCTAGGCCATTTGAATCATCTGCTAATTGGTTTTCC  
TGATTGCCACGGAACTTCTCCATCCCTTCTCACATATCAGCCACAGA  
AGTATCTCCAAAAGCAAATCTGGTGACATGAAGCCCTTGACAAAACCC  
ATTCATTACTGGTTCCACACCTCCTTTGTGGATAAGTTCAAGCTCCTGAG  
TGTGGCAAGCAGGGCCCACCTGGAATCCCTGCCCCTCCTCTCCTATCCCA  
CGCATCAATCTTCTGTCTATTTGCAGTTCCTTGAATGTGATATTCTTT  
CTAGTCTCTGTGCTTTTGCATAACCTGTTCTTCTGACTGGAACTCCTT  
CTCCTCCTTGTAGTTTGGCTAATTTCTAGTCTTTCAAGACTCAGCTCATG  
CTTCACCCCTCTATAACAAGTCTTTCCCAAGCTGGGTGGTGGATGCTC  
CTCTGTGCTGTGTGAGTCTTGAACATCCTCAGCAAACCTCAGCTTTGTTT  
GCTTGTCTCCCTTGCTGTCAATGCACCTGATTCAAGGCTGGCATATACTG  
TTCACCTCCATGACTGGCTCATGGTGGTGTCTCCGTGAATATCATCCACCC  
AAACGGATGAGAGCTACCATGCCATCACTTGTGACTTCCATCTGGAGCTA  
ACCTCCCCCGACAGGAAAGCGTTTCTTAGGAAAGAATATCTTTGGGTTA  
AATAGAAGTAGAGACTCACCAGAAGCACTATGTCCAGCTCAGAATGAAC  
GCTCAGTAAGCAGCCTTGTCAATGAGGAGGCAGCAGGCCAGCCCCAGAGG  
CCTCAAAGTGGGAGAGTAGAGAAGCGCAGTTCTGCCACAAAGGCACAGT  
GGACACCTGCTCCCCTGGCTGGCTGGAAGCAGATGGTGTCCACCTGCTT  
CCATGGGAATTCTGCACCTTTAATAAAGTTTTATGGGACAGGAAGGTGAC  
TGGCATTGACATTGTAACGAGGAATGGGTGGTGGCACCTTTGCTGTGTCT  
TACCAGAAATACCTGTGGCAGGTAAATTTCTAGAGAGACCTCCCATTTT  
TCCCATATAGCAATTTTGAATGTTTCTGAGGGCTTTCCAAATTCATCT  
GGGAACATAGGAGTTCCAGAAAGATGAAATCAAAGGTGATGGTATGCCAA  
AGAAAGTAGCTTTTAGAATGACTTACATTAGCCATTTCATCCATTTCAGCAC

FIG. 3 (28 of 52)

30/118

>Contia35

FIG. 3 (29 of 52)

AGTATGGGAGTTT CAGAGGATAGGGGGTAAATGAGGGGAGTAGGTGGGTAGA  
AAAGGTTAAAAGTAAATAATGATGGGAAGGAAGACAAAAGACGACAGGG  
GTGCCAAAGGACTCTTAACCTCATCTGAACGGAGTTGCCCTGTTTGTCTC  
TCTGATGCTCATGTATCTATCCTTAGAGACAGCTTGGCGGGCAATGTAGA  
GCGTAGGGGCTGACATAGGGGGTGGAGTCCCACCTCCGTGACTTCTAGC  
AAATTAGCAAACCTTTGCTGCTGCTAAGCCTATAAGGCGGACAGAAATGCC  
ATCTTTAAAGCTTGTATGTAAAGTGCCTAGGACCTCGTAGGCATCAACA  
GGAAATAATGGATGAAACAAAACAACGGTGCGTATCTTGGAGAAAGTGGCA  
TCTGAGCAGGAGTATTTTGAAGGTAGGAAAGGGCTCCAAGCACATCTAA  
GAGATTAGGGAACGCAGAAGCCTTAGCCCTGGGTGCAGATTTAACCAATC  
AACTTCTAACCACCGCAGGCTGAGAGGTGTGGAGTGAGAGCCCCGCCAGA  
GGCAGGAGACCCGGGCTTCGGCCAGACCCCGCCTCCTGGTACAGAGGACC  
ACGCCCCGGCTCTGCTGGAGCCAAATGTGGATCAAAACAGCGCGCAGCTT  
CCCCTGCTGGTGAAAACCCGAGCAAGGGGCTCAGTTTCTTTATCCGGA  
ACGTGGTGACAATGACATCTCTTTGCAAGGCTGCTGCAGGGCTTTCTGGA  
AATACGCCCCGTGAGGTATCTGGGCCTGCGCACAGCCTCCCCCGCCAGGA  
CCCAGAGCTTACCTGGGGTCCCCGTCTGCGCTCCCGGATGGAACCGC  
CCAGGGGAAACTTAGGCAGGCGAGCGGACCGGCACCTCCCGCGGACGAA  
CTCACTCGGTGGCCTCCTACTTCCCCGGCGTGTTCACCGCCTGAGAAT  
AACGGGAACAGCGGTCTACTCACCGACAGCGGCAGCAGCGGTAGGCCCCG  
GGCCCCACCATGACTCTTCAGTGACAGTTTTTCTTCAAACGCGCSCCTG  
TAGCCAGGACCGCGCTGCCGCGCGTCCACGCGTCTCATTGGCTCCTGCG  
GGTTTGAAACTCGCTAGTCGTACGACGGGAGGGCGGGACAACAGGCAAT  
AGGCTCTTTGCGGTGGCTCTGGCCTTGAGAACCCGACCTTGGGGCCCTT  
TGATTGGAAGAACGTGCAGCGCACCTCGGCATTGAGGGCGGCTTCCTCGG  
GGCGGGCGCGCCCGCCTCTGAGTGCCTGTGAGTGCCTCCGAGTG  
GGCGTGGGACCTCCGTGGGGGCTCAGCCGGGCTGGTGGTTGGGGGGCG  
GTTACGCTGAATCCAGCTGGGGTTGGCGCGCGGGAGTCCCTGGGCGGAG  
AGACAGGGCGGTCTCTCCAGGATGCTGGGGCGCTACCTGATTCTGTCT  
TTCAAAGTCTCAGACTCACAGGAGCTGTGAAAAATAATATTATAAGAG  
GACATATGGGTCTTATGCATCTAAAGGCTCCTAGTTCTTAGTACTGCAGG  
GTGGCTCGTTTAAATGTGGTAAATATGCATAACATCACATATACCATT  
TAACCATTTTAAAGTGTTAAATTTTCAAAAATGTGCAGTTTAGTGGTAT  
TAAGTACCCTCACATTGTGGCACAGCCACCCTACTGTCTTTCCAGAAC  
TTTTTCATCTTCCCAAATGAAACCCTGTACCCGCTACTAACTCCGCACTC  
CTCCCTCCCCCAGCCCCAGGCAATCACCATTCTAGTTTCTGTCTCTATGG  
ATTTGACAACCTGTAGGTGCCATATAAGTAGAATCATGCAGTATTTGTTCT  
GTGACTGGCTTGTTCCTTAGCATAAAGTATTCAAGGTTTATCCATGTG  
TAGCATGTGTGAGAAATTTCTTTCTTTAAGGGGAATAGCATTTCGTT  
GTGTGGAGATGCCACATTTTGCTTCTTGGTCCATCCCTCTCCGGACACTT  
GAGTTGCTTCCACTTTTTGGCTATTGTGAATAATAATATGAACATGAATG  
CACAAATAACTCTTTGAGACTCTCCTTTTCATTCTTTTGGGTATATACCA  
CGAAGTGGTATTGTTGGATCAAACGGCAATTCATTTTTAATTTTTTGAG  
AACTGCCTTACTCCTCTCACGGTGATCTCTTGTTCAGGTATATTTTCG  
ATTTACGCTGATCAGCTGACTATAAGGCCATAAGGCTAACGGAGAAACGC  
AGGCCTAGTTTCTCTAGTTACTAGGAGATCGCAGGCCTCGTTGTCCTGA  
ATCCCTAGACACACTTCATTCCCCCTGTTTTAATCCTAAATTTTTTTCT  
TTTGAAGTTTGTCTGTTTCATCTATTCTCCAGTTTCTTAAAGAGGTCTG  
GAAAATGCTTTTGGCTCCTTGTGTATGAAGGTTCTCTTCCATGGATGCT  
GGAGAAGTCGTGTGTGGAGGGGCAGTCATATCTGGGCACCTGTTGGCCAG  
GTTACAGCTTACCAGTTGGGTACTCAGCAGGGCATGAAGCCACTGCAGCAG  
CCCTTCTCTTTAGCCGTAAATAGGGAGTTTGAAGAGAGCCAGGGTTTCT  
GGATTTATGCATTTTGATATTTTCAATAGTGTATTAATGTTTAAATAG  
GAAACTGATCATTATTTTGTAAATGACTGAGAAAGGGACTCCTTCACC  
AACAGTTTCAGAAAAGTGAAGGCGTTTTGTTTTGGTCTTTGTAGAATCT  
AGGTGGTTGAATGCATGTGAGTTGTAGAAGTCACCTTGCCTGATATCCCA  
CGCAGTGCTGGAGTATTCACAGACCCCATGTAGGTACTGCACCTTTGCA  
GGTATACTGCTGGTGTGGTGAGCTGCCTTACCTGTCTGTTATTGGAGA  
CCCCTGCTTATTAGGAACTTAAATGAACTCAAATGAGCTTCTTGCTT  
ACTGGTCTAGTCCTTTGGAGCAACATAGGCCAGTTCTGCCTCGTTTTTT

TCCATCCTTTGGGTATTTGACGGTCTATTTTGTAGGACACAAAATGTGGG  
AAAATAGCTAGGCAGGTTTAAAAATTCTCAACTCTACCAAGCATGGTGGC  
TTATGTCTGTAATCAATCCCAGCACTTTGTGAAGCTGAGGCAAGAGGATT  
GCTTGAGCCTAGGAGTTTGAGACCAGACTGGGCAACATAGCAAGACCTCG  
TTTCTTAAAAAATAAATAATTACAAAAATTAACCAGGCATGGTGGCA  
CACACCTGTAGTCCCTTCTACTCAGGAGGCTGAGGTGGGAGGATCACTTG  
AGCCCAAAGTTGAAGGATGCAGTGCAGTGTGGTCAATGCCACCGCACTCC  
AGCATGGGAGGCAGAGCAAGACCCTGTCTCCAAATAAATACATAAATTAA  
ATTCTTAATCATTCATCAAAGTATCCACTGTAGCTTTCCATCATCCTGG  
TGTGTTTTTTTTTAGAAGGATCTGGCTCCATTGCCCGGCTAGAGTGCAGT  
GGCATGATCTCAGCTCACTGCAGCCCCACCTCTCTGGCTTAAGCGATCA  
CCCCTTCAGTCACCCCTCTGGGTAATTTTTGTATTTTTTGTAGAGATGG  
GGTTTTGCCATGTTGCCCCAGGTTGGTCTTGAACCTCTGGCTCAAGCGAT  
CCATCTGCCTCCATCTCCTAAAGTGTGGGATTACAGGTGTGAGCCACCA  
CACCAGGACAATCCTGGTGGCTTTTAACGGTTTTCCATTGCTCTCAGGCT  
AATGACCTATAAGCCCCCTGCGGGCTTGGCCTTTTACTCCCTEAGCATTAG  
CCACCTCCCTTAGCCTTAGCCCACTACTCTCCCTTGCTCAGTGTAT  
CCAGACACTTTGTTTTTCTTTCCATACTCCTCTCTGTCTGGGAATCCA  
ACCTTTCTTTCTCATTTCTCTAGTTGATTATTATTATTTTACTCTAGCA  
GCCTTATTGAGATATTTACATACCGTACGATTCTCCCACTTACAGTGTAC  
AATTCAATTTTCTAACATTTTTCATCACCCCTTAAAGAAACCTATACTCA  
TTAGCAGTCACTCCCCATTCTCCCTCCTCTCAGCCCTAGAAACCATGA  
ATCTACTATCCATCTCTATAGATTTGCCTTCTGGACATTTTCATATGTATG  
AAATTATGCAATTTGTGGTCTCTGATGGGCTTCTTTTGTACCAAATAT  
CATGGGTTTGTATCTAGGTCTCTGCTCGCTGCACAGAAAGCCAGCCACT  
GAGATGACAAGTATTGCCAAGGAAGAAGGCTTTAGTCAGGTGCTGCAGCT  
GAGGAGATGGGGGCTCAATCTCAAATCCATCTCGCTGACCTAAAACAGG  
GGTTTGGATGACAGGGAAGAAATGTAACAATGCGTAAGAAAACAGGAACC  
AGGGAGGGGCAAGCAATCCTGATGAATGAGTGGTCCAAAGTCTCAT  
TGCCTGGATGTGGTGATCTGGCGAGTTTCAGTTCTTTGATACTTTTTTTG  
AGAGGCCTGAAGTCTTTTCCCAGGAAGGAACCTCAAACAAAACAAATACA  
AGCTTCCAGCTTTAAGACCAGAAGCGTCAATTTCTATGTTTATCCGAAAG  
AACAGTCTATGGGACTATTGGTTAAGTTTCACTTTCACTTAGTATGCTGT  
TTTCAAGGTTTATCCACATAGCATGTGTGCTAGTACTTCATTCTTTTATGAC  
TGGGTATTCTATTGTGCGGATATACAATATTTTATTTGCCATTTCATCAGT  
TGATGGACATCTAGGTTCTTTCCACTTTTGGCTATTATGAATAATGCTG  
TTATGAACCTTTTCATGTATAAGTTTTTGTGTAGACATATGTTTTCAACACT  
CATGGGTATATACCTAATGAGAGGAATTACTGTGTACATACGATAATTCTA  
TCTTTAACCATTTGAGGAAGTCCAGACTGTTTTTCAAAGCAGCTGCAGC  
ATTTTACATTCTTACCAGCAGTGTATGAAAGTTCCAGTTTCTTTACATCC  
TCAACAACACTTGTATTGTCCATCTTTTAAATTACAACCATCCTAGTGG  
TTGTGAAATGGTATCATTGTGGTTTTTATTTGTATTTCTTGATGACT  
AATGATGTTAAGCATCTTTTTATGTGTTTACTGGCCATTTGTATATCTCT  
ATTCAGAGTCTTTGCCAATTTTTAAATTGGGTGAGTTGTCTTCTTCTTT  
TTTTTTGAGATGGAGCCTCACTCTGTTTCCCAGCTGGAATACAGTGGTGT  
GATCTCAGCTCACTGCAACTTCCACCTCCTGTGTTCAAGTGATTCTGGTG  
CCTCAGCCTCCCAAGTAGCTGGGATTACACGCACCTGCCACCATTCAG  
CTAATTTTTTTCTTTGTATTTTGTAGTAGAGCGGGTTTACCATGTTGG  
CCAGGCTAGTCTCTTTGTGACTCTTAACCATCCTTCAGTCTCAGACAAA  
ACATCCCTTTCTCAAGGATTGTGATTAGCTTGATTATTTGCTTATCTTTC  
TCCCTGCTAGTCTGTAACTGAGGGTAGGCCACTATATTCATTGTTCTTG  
GCACCAAATAGAACTAAATTAATGTCTTTTGAATGAATAGGGCTTTCTC  
CTTTTAAAGATCCCTTCAATACAGTAACCACTATATATAAGTAGCCAC  
AAGCCCATCAATAATACTACTAGTNTCTTGGCCAAACC  
>Contig36  
GGCTCAGCGTTACTATACTGGTCTCAAACCTCCTGGGCTCAAGCGATCTGC  
CCCCCTCGGCTTCCCAAAGTGTGGGATTATAGGCGTGAGCCACGGTGCC  
TGGCCTCAAATAACTATTTAAGTGAAACAAAACCTAGTATGGCACTAATGA  
AAAATGTATAAATCCATAATCGCAGAGGGATTTCAACTTACTTCTTTCTGA  
TTATGTAAAGGTCAAACAGACAAAAGACAATGACAAAACCTAATGCAATG

AACACTTTTTGATTTAATGAACATATATTGGATATGTACCCAAGAATTAGA  
GAATACATACTAGTTTTGAGTTTATGCAGAACATTTACAAAAATTTAGTG  
GAAGCCTAAATTATAAAAAAGTTGCTGTCACGTAGAATAACACACAAACCC  
CTGAGTCCGGAATTCAAAGCCCTCCACACTCTCCTCTACCTTTGCATCTT  
TATCCTCCACCACACTGCAGTGCATACTCTGGGCTACTACTCACTGTTCT  
TGATTCAAATTCATGTTCTGTCTGCTCAAATCATTCTCTCTGCCTGGAA  
TAAGTACTTTCATACATATTCTGCTATTGAATTCTTGTCTTAGCACCCCAT  
CTACTCCAAGACGATGTCCAGTTGGGGTTACTCCCTGTCCCATTCTTT  
GATTACACTTTTTTTTTTCTACTTCCATTATATTATTGATCACATCTGTGC  
CACAGTTTTTTGACTTTGTGTCTGCTTTTACTCTTTTCTAGACCCTGATAG  
CTCCTGAAGGGTTGGGTCAATTTCTTTTTTATTGCTCATTCTCATGGCA  
CAGTGAGTGCTTAATAAATGGCTATTGACTGAAATTAACTGTATCTAAA  
TGGACATATTCCACTTCTGGGCCATTCAATCTTTCTTTCTATTGGAACCA  
GGAGATGGGGAACCATAACAAAGGTAAGGTTGTGCCATGTGAAAGAACAT  
GGAACCTTCCCCTGAGGGGCCAAAAAGAGCAGGGAAAGGTGCAAAGACAA  
AATCTTCCATTTTTAAACAATGTAAGAATGTGGTCCACCTCATGCTCAGG  
TGGGACTTTATCATGACGTTATTTTTGGGGACTTATAGCTGCATCATTTA  
CCCCATATACATTTACCTTTAGTGTAGGGAAGTGAAGGACAGGAATTTGT  
TGATGCAGACTCTTGCTAATGAGGCTAACACTTGGAGAATTTTTATCATG  
CATTCAAGAAGCTTGTTTTACATTTCTTCATTAATACTTTAGTTGGTGGT  
TTAGCTTTAGTTGTAGGCTTATCAGATATTTGGAGATATCTTCATAAACG  
ATGGCTTTGGTTTTAGAAGAGTTATTCTGAAGCTACTATTTCTGGCAATA  
ATCAAACAGCATGGCCATTTGTTTTGTAAGGCCTTCTCTAGAATATGACG  
GTAAATCTACGTGTGGAAAAATGCTTATTCTTCTGTCTCTATAAATGT  
GAATCTAGTTTGTCTTCAAATGAAATCAAGTGATTAAATGTAGTTTTT  
TAAGAAGATAAATGGAGCAAAGCACTCTGTGTTTACAGTGTGGAAATC  
ACTCATCCCTCATAAACTGTCCCAACTGATCCTGACTCACATGAATGAA  
TTAAATAAGAGTTAATAACATCAATTTACATTTTTTAAAGACACTTTCCC  
ATGTTTTAGACTATTGGTTGGAAAAGCTGGTAGGTGTACAATTTGTGGAG  
AGTTGGCTGTTTTGTCTGTCTGTTGTTGACGTATTTCAAAGCCATATCT  
AATTTTGTGTCAGAAATGGTCTGAATTTCTACAAAATGTTGAGTTGTGTAG  
TGTGGAGAAGTACGGAGCCATTTACTGAAAGGCTGGGGGGAAATGACGAG  
ACCCTGAGATAAGGCAGTAGTGGTGCGAACAGAGTGGAAAGGGAGGTAGTT  
GAGATATGTTGAGAGTAGAATCAGAATGGACATAGTGAACAACTGGATGC  
AGGTGGGGGCTGAGGAAGCAAAGTTGAGGATAATTCTGAGACTTCTAGGT  
TGATCCACTGAAGTTACATTATTCAACACCACAAGGAACTAGGGGAATG  
AGAAGGCATACTGGTTTGTCTTTGGAGTGGAAAGGGCAGTGATGTAAGAGGA  
GTTAATGAGTTAAAGTTTGGATATGCCTGAACTTCAATTTGATATGTGCA  
TCTGATATACCCCTGGGGTGACCCCTCCAGGCAATGTTGAACATGTGTAT  
TTCTTAGTAACTGATAGGCATCACAGACTCACATCAGTAAGGAAGCAACA  
GCAAACCTTGATTGGACGATATACCTGGAACCTCAGTACCCTATGACTGGAG  
CAAGTCTCTGTCACTGAAATGAGGATAAGAAGAATCTTGACCTTGTTGGAA  
TATGTTGTTAGGAATATATGTGATGAACAACATAGGATACTTCTACAGG  
GCTCCACATGTAAGAGGCTTTATAAATGCTTGATAAATATTATTGTTG  
TAATTTATTTTCAAAGTAAGATGCCACTGGAGGAATCTTTGGAACCCAAA  
TTAATAACAAATAGGACTGGATGCAATGGCTCACACCTGTAATCCCAGCA  
CTTTGGAAGGCCAAGGCAGGAGGATCTCTTGAGCCCAGAAATCAAGACC  
AGCCTGGGTGACACAGGGAGACCTTGATCTATGAAGAATTAAAAAAAT  
TAACCAGATGTGGTGGTGACGCCTATAGTCCCTGCTGCTTGAGAGGCTG  
AGGTGGGAGGATTGCTTGAGCCCATGAGGTTGAGGCTGCAGTGAGCCATA  
ATTGTGCCACCACACTCCAGACTGGGTGACAGAGTGAGACCCTATCTCAA  
ATAAATAAATAAATAAATAAATAAAGTACAAACCAGCAAACTAAT  
CCTTTCTAGAGATTATTGAACTCTGGAGGGCAGATCTGAATGGAGCCAGC  
AGAGGGACCTATGGAGATCAGCCTGGCCCTGGACAGCACCAGGCAATGGG  
GTTGCTAGAGAGGTAATGGGGTTGAACAGGGTTTAAGCCATGAGGTCTCA  
AGAATCCGTGAAGACTCAGACTAATTTTTTTTTTTTTTGCATGAGGATTAG  
GTGTTCCTAGGAATTTCAATGAGAGCAGGGTTAATGAAGGAATGCAGGGT  
AGGAGAGCTGAGGGAAGGCATCTGAGAGAGCCTGGCTTATGAATGGCTGC  
CTCAGTATGGCTCACCTGCTTTCTTGTATCTACTTAGCAGATGATCCCA  
CCCCAGGCCCTCAGGGCCAAGGTCATTTCCACATAGTCATGGGCCCTTGA

FIG. 3 (32 of 52)

34/118

GGGCCTGGAGCAGTGTAAGGAAGACAGAGTCTTAAGAAATTGCATTAAAC.  
GTCAATGGTGTCTGGCAAGTGTCTGTCATCCTATGCCAAGCCTGATCTGAAG  
GGGTGCATGCTCATAGGTAGCTGCTGCCAAGATTACAGCAGCTTCTTCA  
ATCCCAGATCCATGCTCTCCTATATTCAATTTTCCAGGGGTTCTGTCTCT  
TCGACAGTGATGAGATGCAGAATGACTTATTGAGTTATTCTCCTGATAGT  
TGCCAACTTTTCCAAATGACAATGGGGCATGGAGCTTGAGAGTGGAATG  
AGGCCCTAGGGATAGCGTGTCTTAGGAAAACACTCCCAGCCTGATGTAATT  
CTGGGGGTACAATGGCATTTCATCATCAAGACTGATGTAAAGGGTGACT  
AGCAGTGAGTTGGGGGTGACTCGCACTGGGGCTAGGTTTCTGATTCCTGCC  
TAATCCAGACAGAGCAGAAGCACTAGTGGGCTGGTAGAGGGCCTCCAGGG  
CCTCACTTAATGTCTTGGAAAAACAGCTCCAGATTGTTGGTTCACGTTCT  
GAGGACAAGCTTGGGTACTACAGGATAGAGAGAGTGGTGGGAGATGCCGT  
GGCCTGCCCTGTGATGCCTGCCCTGCCATTCTGCGTGTGATGTCTCTG  
GGGCATCTTGCTTCCCTGCCCAGACCTGTAGTTCAGCTGAGGGCATGTG  
GAGGCCAAATGGCTTCTTAGAGTGTTACTTTCCTTGAACAGCTCTGCTGG  
GAGAACTGGAGGAGCTAGCTAGTCACGGTAACTGCAGCAGTCAAAGGATC  
GTCCCCGTGGAGGTGGGGTGGAAAGGTAGAGAAAGAGAACATATAGCGTT  
TTCCCTTGGAGATGTGTGGGCATGTCTAGAGGAAATACCCAATTCCTGAG  
CCTTGAGCCCTCCAGGAAACCTTGGAAATATTAGGTTAGTCATCCCCAAGG  
AAGTCTAAGAAATCTGGTCTCACCCATCTCCTTTAATTCACCAATGATC  
CTACATGATATTAAGGAACACGGGCCAGTAACCCCTCCAAGCAATGGATGT  
GGTGGTGAAGTTTGACCTCATGATGGAGCGGAGGTTGGTTTGAACCTTAA  
GAATTTAATTTATTGTTTCAAACCTGTTCTCCACTCAGCGTTATTAAAGCA  
TACATAATTGACACATAAAAAATTGTATATGTCTACGGTGTACAATGTGAT  
GTTTCGATCTATGTATACATTGTGAAATGATTACAACAAGCTAAATAACA  
TACCCATTCTATCGTGTTCAAAGGAATTAACTCAAGCACAAAAGAGAGG  
TGCTGTTGAAGAGTAGGGCTGCTCTATCTAAGTAGTATGTCTGGGGTTGT  
CCTGGATCAGGGTCTTTTTGTGCTAGTAATAAACCAGCCCTTCTGGGGCT  
GCTCCACTTTCCTCCACATTTTCTTCTGGAGCCTCCCTAAGAATTAGGACA  
TGGCCACTTTCCTGCTAGGCTTCTTACTTCAACAAGGACAGGGCTTGT  
GCTGCCCATGCCACTTGAGTGTCCCTACAGCACAGAGCTGAGTGCACAC  
TGGCTGAGTGAGGAAATCCCCAGATTAATCTTGGTTCTAAGCATCATGG  
CTGTATTTACACGTATATGAATTACAAATTACAGCATAGTCAATAAGG  
ATTTTTGTGCTACAACCTGGAATCCCAGATTATGCAAATTGGATAGTATAA  
TATTGAAATTCCTAGGACTTTTTATTAGTTTTAAAAAATTATACAAGCTT  
AGAGTAAGAAATTAAACAGTGCAAAAGAATTCAGTGTGAAAAGTAAATG  
CTCTGTCTCTGCTGAGAGACAGATATTGCAGCCCAGATACTACTGGGGTC  
AATAGTTTTCTTTAAGCATGCCATTTTGATGGTTTATGGGACTTACAGCT  
CAAGAAGCTTGACACTAGGGTTGATCTCAGAAAATCATTGTTGCAGGTAT  
TAGATTAACCGTCTCATAAAGATACACACACAGACACAGCGATTGGAGA  
TATTCACTGGGGCTTATGGGCTGCTTGTCTTTCTGCTCTGTGCCTAAGT  
TGGGCTCAGAGTAGCCTGGCATCGGCTGTGGGGAGAATGCTGGCATGGGG  
TTAGCAGGAGCCCACTTAACATGTCCTAAGCCACCTGGAAGAGTCTTCA  
AGGAGACCAGACTCCAGAGGCCCTAAGGAAGGAAGGACTTTTGCCCCGTTT  
TTAGGTATTCTAGTCCAGAGTTTAGGGAGGAATGGTTTGGCTTTGGGTC  
GTGTGCCCTTTTACCGAGTGGGATGGGATGTGCCCATGAGCTGTTGAGCT  
GGCTCTTGAGAAGACAGCAAAAGCGGGAATAAGAGGTGAGGAAGCTGTG  
TGGTTGTAGGAAATCCCAGCAGAGGGCCTGGGGGTCAAAGTGGTCATGG  
TAGTGACGGTGGAGGCTGAGGTGGTAGAAAATCAGAGGACAAACCCCATG  
GGCTGCTGGTGATCTGACCGAGCTCCTATGCTCTCCTGGTTTCAATTTAGG  
CTCTGTAGCAGCAGATGATTGGCTGGTGTGAGAGCAGTGCACCTGCCATA  
TCAGGCAATCCAAGACAAGTCCAAGCTACGCTGGGAGGAAACCTGAAGGC  
AGCAGCAGGTAGACTGGCTGAAGACAGACAGGCAGGCAACTTGTCAATCA  
GATTTGTGTTTTTAAGGACTTTTAACTGGGGAGCCCTCCGGGACAGATCA  
GATGAGAGTGAAATGTGCTCCGCCTTAGCC  
>Contig37  
GGCCGTTTCGCAATTCTGTAAAAGGGAGAGTGGTTTTATTTATTTTAAAC  
ATAGTCAAGCTGCTAAAGTATATGATATGTATAGATAGAGTATAATTAAA  
TACTTTCAACTACAGACAAAATCAGGAGAATGGAATTAAAAAACAATTTA  
CAAATGGGTAATGGCAGCATTGGGTTGCGCCACCCACGAGAAGGCAGAC

ACCAAGATTCTAAGATCAACGTGGCCAGCACTTCAGACTTCAAATAGAA  
TTCGTGATTATGCATTATTTTTCTCGGAAAGTTTTCACTTCACTATATGC  
TACTTGACACTTGCTTTCTCCTAAGACATCCCTCTATTTTTGAGATGACTAA  
CTCAGCAATTTCATTTCTCTCACGCATAAGCTGTCACTCAACCCAAACCCA  
CCAAGCCTGCATTCTACCCCTCAATAAGGTCTTGCTGTGTAACTGACCCA  
CTTCACCTAGTTTCTTAGCCCTCTCTTGACCAGACATGACTCTTTCATAA  
GCTAGACCTATAAAGTCAGGGCTCTTAAGTAGCTGATCTCTGATAGTGCC  
AAGTGTCCCCCACTGTTACATTTTCCACTCCAGCTTCTAACAGGTGATA  
GACTGCTTTTTGGGGGTAGGGGCACCAAAACATATAGACCTCATGTTTGG  
ATGTAGACACTCCAGTTTCTTTAAATTACAACCTACATATTAATAATGACT  
TCCAAGTGTACATTTCACTCCAGATCTCTCCCTGGATCCCCAACTTTGT  
AAAACCCACCGCTAGTTGATATCTTTTGATGTCTGACAGGCATTTCAAA  
TTAATACTGTCAAAACAAAGTTATTGATTTTCATCTCTGCATCTGTTA  
CAAATTTTTCTTACTTTGGTAAATAGCACCCAGGCTGTGTCACTGCCAA  
GAATTTCCACAGCTCTTGGAAATAAAATTCAAATATTTTTCCAAGGCAGA  
AAGGCACAGTGTAATCTGGCTCCTGCCTACCTCTCCAACCTCGTATCACA  
CTAGTCTCCCTGTCACTCACCCCTCCAGGAGCTCAGGTATCCTTAAAGT  
TTCTTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAACAGTTTTGCTCTGTT  
GCCCCAGGCTGGAGTGAAGTGGCATGATCTCAGGTCACTGCAACCTCCGCC  
TCTTGGGTTCAGTGATTCTTGTGCCTCAGCCTCCCAAGTAGCTGCAATT  
ACAGGCGCGTGCCACCACACCGGGCTAATTTTTGTATTTTAGTAGAGAT  
GGGGTTTTCAAAATGTTGGCTAAACCGGTCTCAAACCTCCTGACCTCAAGTG  
ATCTGACCACTTCAAGCCTCCCAAGGTGCTGGGATTACAGGCGTGAACCAT  
TGTACCCTGCCTCCTTGAAGTTTCTTGATCCAGACTCATTCTGCCTTAA  
GGTCTTGCATCTTCAGTCTCTCCCTCAAATGACACCTCCATGAAGACGCA  
ATTACCTGTAATTACCGTGTCTTATTTAGTCAATGTGTTGGTTTTCTGTC  
TCCTCCACTACAGTGTAAGCTCTATGAAGGCAGAAACCTTGGCAGTCCAG  
TTCCCAGCACAGTGCCTAGCACACATAGGTATTTAATAACACACAGTAAA  
ATTCACCTTTTAGTGTGCAATTCTGAGTTTTGACAAATGCATCAAGTCAT  
TTAAGTCTGACTATTATCAAGCTATAAGATGGTTGCAACACTATCACTAA  
TTCCCTCATGCTCTTGGTAGTCAGTCTCACCCCTAACGCCCCCTCCTG  
GCCAATCACTGATCCGTTTTTTGTCTTTATAGTTTTGGTTTTTCCAGAATG  
CCAATAACTAAGTTTTGAATGAATGAATGCTATTAACCTCTCATTTCTGAC  
TCCAGAGCAACATCCATGCAATATTTATTATTTAGCCCCAAATACTGCC  
CCCTCACCTTCACTCCAACCACCTACTTGATGATACAAGGTGAGACATTT  
GGCATGTGCTTCCCTCCATGTTCCCTAGCATTTTCCCTATCTCCTTAGCCTT  
CCTTCTAATCATAAACGAAGAGTGAACCTTCCCTTTCTAAAGGCAACTTA  
CTCCTAGGACCTCGATGCCATAATTTTGTCTCTAGTACTTTCTATATA  
TACACCAACAATTAGCTCCAGAAAGGTAAAGACTCACTGTGTGCTCATC  
ACTGTGCTCTCTAGCGCCTGGCACACTGCAGGTGCTGAAGAAACACCTAC  
AGAATGAGTGAATGAATCTCTCCCTCTCTAGACTCCTTCTTTTTGTAAT  
CAAACATGTTCAACCTGCAACACAGTCTTATGACCAATCCTCTGTTGTCT  
GACCTAGGCTGAGCTCCAGGGCTGGGACCCTGACTTCCTTATTCACCACC  
TCAAGGTCTCTGCACTCACTTCTCTTTCTGCTCAGGATTGTTTTCTTCT  
TGTCACCAGTCTTTTCTCAGACTTAGGTCTCAGCTCAGACATTGCTGTTG  
AAAGTACTTCTACTGATCCTTTTATCTAAAGCAGCCATTCCAGCCCTACT  
CTCTTGATCATAGCACCTGAATTAAGTTGTTTACTTACTGTCTCTTCAG  
GAGGGCAAGGAGCTTGGTGGTGGTGTTCAGGGCTGTACCAAGCTGTACCT  
TGCTTCAACCTGCTACACTTTTTAGCAACCATCTAATTTTACATGCTCCC  
TTCACCTCGTCAGAAATTTCTTATTTTCTACTTCAAGCAGGTATACATAT  
GTGCTTCTCCTGGGAGGCTCACCCACTTCATGAGACTACATTTGGTCCTG  
GGTAGAAAGTGTACAAATCCACTGGCTCAGTTTAAATCAATGTATGTTA  
ATATTAACCAACCTGAGATCTTGATTTCCACGCCTGGCTAATTTTGTATT  
TTTAGTAAAAACAGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTCC  
CGACCTCAGGTGATCCGCTCACCTCGGCCTCCCAAAGTGCTGGGACTACA  
GGCATGAGCCAGCGTGCCCGGCCTAAGATCTTGATTTCTACCATCTGAAC  
TCTGTATTTGAACTGACTGCTCCTGCTTGAGCTTACTGGCCAAACTTGG  
CCCACTCAGACTCACGGAAGTTTCTGGTTCTTCCCTGGTAACTTTTCTGA  
ACTTAACCACTGGTTTGCTTGACAAGAGATTACCATCTTCTCACTTCCTA  
GCTATGTGAACCTCACTTATCTGCTCTATTGCTGTTCAGTCTAGCACGGCA

FIG. 3 (34 of 52)

36/118

CTTATTGAACGAGTGTCTACATCTGCACCCCTACTTCTTACTCATCCAT  
TCTGTTTCAATTTCTTAAAAAGAAAAAAGCTATTGTAAACATACG  
ATTACAGAAAATGATTTATAACATGTGTATGTACCACCTAGCCCTGTCAA  
GTCTTAATATTTGTTATATTTGCTTCAAATCTTTTTTTCAGACTGTAGTTA  
AAAATTACTTAGGAGCCATTATTTATGGCCTATTTCTGACCTAGTCTTC  
TTGATGGTCAATTTGCCTAATCATCTTAAGTTGCAAAAGCTTAGAATTAA  
AGCAAAGTACCTTCGATCCTCTGCTGTTGCCTTCTTTTTAATATTTGGGT  
TTGTTTGGGTCCCATTACGGTTGTGACATCAGCTTGAGTTTGGGAGCT  
GTCTTGTTCAGAAAATGGTCTGCGGGAACAGCCTTTTTCAACTTGGAGTC  
CAAAGTCTGTGCTTTTTTGTGAAAGCCATTATTGTTATGTTTATTACCAC  
TGGTTCATTTGGTCTTATGCTAGGGGTGCTTGAATGGCTGAATTAAAT  
CTGCCAAGTGTCAAATTAGGCCTCTGGCTTACGGCTTTTGACTTTTGCAG  
TACACATGATGTCTGAGGTATACAACTTGGCTGGACTTCTGATCTTGCT  
TGATGTTTGGATGTCTGTTGTTATATTACCCTGAAGCAAAGTGGGTAT  
GTTCTGGGTTTGGTGTGCTTCACTCTCTGTTCACTAACAGGGTATGACCG  
TATCTTAGTTTCATTTGGTCTTTTCAATTTGACTCTTATTAACCTTTATAT  
CTTTGATGTTCTTGAAGTACTGGTTTCTTTGATGACTGAAGTTTACTAAGG  
GTCCGAATAAAGTGAGAGGGAACCGTCTTGGGGTTTTACTCCTGGTCT  
TGCAAGTCTGCTCTCTCTAGAGAGTTGCTGTGATTTTACTGGGAAAGTCC  
TGCTTTGTGTTTTCTCAACAAATTGTTTATTAACCTTCTTTTTCAGAAC  
GCACTATTAAGTGAAGTTTGGCCCAAGGCTTGTGTTAGGAAGTAAAGTGT  
CTTGGTTTTGATTATAAGAGTCACTCTTTGGCTTACTTCTGGTATATAATT  
TAGGATCTGGCTTCTCTCAGGTTCTGTTAAGATATCTAGCAAGTCTCT  
TTGTTTGTCTCTTTTAGAAAGTTATCCAAAGATTCGTTTTCAACATGGAT  
ATTATTCATAAAGTCTATACATTTACCATTTCTTGATCTGTTAACTGCT  
GCTTTGTAGTTTTCAATTGCTCTATATTAAGTGACCCACAGGTTTTCTT  
GACAGTCTCTCTGTGGTGGACTATCTAGCTTCACTGTTGAAAAGTCTT  
GCTGAAAAGCTTAGACTATGGGTAGAAAGAACACATTTTGAAGTCCGCC  
TTTTGCCCAGAAGTTTTGGTGGCTCTAAGTCTAGCTTCTGGGACCTGCA  
GTATTAGGTGGTCTGGGCTGGAGTTTAAAGCTGATGGACCTTTTAGGTTT  
GACAGGCAAAACAACATGGTTGGTAACATCATTTTTGGGTCTAATAGTCT  
GAAAAACAAGAAAATACATATTAATAAATCCTTAACATATCTTATTGT  
TTTTAAATAAATAACTGTGTTTAAACATGCTAAAAAATAATCATTTTT  
AGAATTTTCACTAAGAAAGTTGAATCCTCAGAAAGTAAAGAAAGACTCAC  
TAATAGGTAGTTTTTGTGTTTTTTTTTTTTTTTTTTTTTTTGGACAGGATC  
TTGCTCTGTCAACCCAGTCTGGTGTGAGTATGCAATCTTGGCTCATTGC  
AACCTCTGCCTCTGGGTTGAAGCAATCTCCCAACCCCAACCTCGCAAGT  
GGCTGGACTACAGGCGCATGTCACTACACCTGGCTACTTTTTTGTATTTT  
TAGTAAAGTTGGGGTTTACCATATTGGCCAGGTTGGTCTTGAAATCCTG  
ACCTCCAGTGTACACGCACCTTGGCCTCCCAAGTGTGGGATAACAGG  
TATGAGCCACCACACCTGTCTAACAGGTAGTTTTTACAAGTGTGTTCC  
TATCAGAAGTATATTAGAATCTTTTAGCTTGACAGAATTAAGCAGAGATG  
CAGTGAATATACAAAAGTCTTCTTCAAAAATGAATTTGCCTCAAACAG  
TAGTTGTTGAATGCCTATTATATCCTAAGTGCCCTCCAAAGAACCTGAA  
AAAATACATACATAATGAAGTTATGTTAGGGTACCTCCCAACAAATCTCT  
CTTAGTACTTTGTATAGCCACACTATATGTTTTTAAACCACTGCCTTTG  
TAAACATCAGATATCACTCAAGAACCTCTGTCTCATCCCTGGAGATCAG  
TGACAAGGAGATAGGTGGCAGATGATGTGAGGCCTGAGATATGCTGCCAC  
AGCTCTCAATAAACATGTAACATCTTAATAGTCATATTTGTAAATCAGC  
CAGGACAGGGTTTTAAGGTTAGAGTCTATGTTAATAATAACAAATGTTT  
AGTCATGTGATTTAAGTTTGGATAAGAAAGGTAGGACTCGATTACAGAGA  
ATTTTGAAAAGTAGGGAAGGGAGTTTAGAATTCATATGGTAAGTAATTGG  
GCAAGCCACTATGAATTCCTGAGCATCTCTCATGAAAGCAATTACTCAGA  
AAGGAGAATTTACAGAGATTTATGGAATATGTTTCCAGGGTAAGATATG  
GGAATGCTAGAGTTACCACTCTATTTTTGATTTGACAAATATTGTGAAGA  
ATCACTACATAAGACTTGGCGAGTATGTAAAGGATTTCTAACCAGAACCAT  
TTGGCATTGAGGGCAAAGAAATGTCTACTCTGGATGATAGCGGTGTGTGT  
GGTGTACTAGGAGTGAAACAGCGGAGTTGGGAGTGGGAGGCAGAGAGAT  
GGATGGTATACCCACAATGGCTATATCTGGATTAATCTTTGAGCACCAC  
ATTTATATACACCTCGGATCTCTCCATCATTGCTTACTGAAGAGGTGGAG

FIG. 3 (35 of 52)

37/118

GGACGTTGGCATGAAAGCCTCCAAATGTGTTTTTTTAGTTGCTTTCTTA  
ATATTAAAAACGAATTGATATAATCCACAAACCATAAAATTCACCATTTT  
AGTAAGTGCACACTTCTGTGGATTTTAGTATAGCCACACTATTATACAGC  
AATCACCACGTGTCTAATTCCAGAACATATTCATCACCCTTAGAAAGAGAC  
TTGGGTTTACTTGTGGCAGTCCCTCCCCA  
>Contig38  
GGTCTACATGTGCTCGCAAGATTGGATATTGAAATATCAGCAAGAAATTA  
AATGACATAGTAGTCATTATGCCTAAATTATTGTTATTTTTTGATTGAAA  
AAAGTTGAATATTTCAAATATCAAGGTAGTAGTGAGATATAATAAAGAGA  
GAGTCAGTTCTAAGTATAGAATTGCTGATTCAAGCTCTGTTCTCCA  
ACATTTGGGCCACATTGAAGAGACCATGTAGCTGCTTTCAGCCTCGGTTT  
CCTCCTTTGCAAAATGGGGATTACACTACCTGCCTCACAGAGATGTAAAC  
TTATGACATGTTATCATGATTGCCAGGGCCACCTGTTTTCTTTTAAACA  
TTGAAATCACTGTGCCTGAAACAGGGATTTCCCTGCCCTTTGTGCAAGCT  
CCAGAAACAGGAGTCAGCCTGAGTCCCGCAGCTAAGAACGTGGATTCTGG  
TCATTTTCTCATAGCGAACACACTTCACAGGTCTTCAAGGGAGTACATT  
TTCCTATAACTCACCTTAATCTCAGTTGAAGCCTCGTTTCTTATTTTGCA  
CTGTGGCCAAAACTAAATCTCATTTCCTTTCACGTAACTTCAGCAATTC  
AATAATAGTACAGTCATTTTATGTTTTCACTGAACCAAGTCAGGGTTCCA  
CTCCTGCCTCCCTTTCTGCTCTGAGGACATCCATGAAGTGGAGGGGGTCT  
TATGTAGCCTGGAGCTATTGGTGGAGGGGCGATGGGTCCGTGGTGGTCTTG  
GGGAACTGCGGGGCTGTGTCTGGCTGGTCTGGTGTCTGGTGATTGGCCTT  
GTTCCACGCGGTTACGCTGCAGGACAGTTCGTGTCCTTCTTGTCTAAT  
GATCAGCTTTTAGGCTCACGGGCTGTCTCTGCTGAGATATGGAATAGGA  
CAGCCTCTGGATCTTCTTTAACTCTCCTGGGGCCACAGGGGACTCTGTT  
TGTGTCTGTGCCACATAGGATGATTCTGCCCAGACCTTTGCTGCCATTT  
CTTGTCTGTTCTGTTTTTAGTCTCTGGAGGGCTTGCAAGTTTCTTTGGG  
GTCCCTGTGGAAGCAAAGCAAGTCTCTCCACGCTCAGATGTCTAAACG  
TATCTGGGTTTTATCGTCCACCCATCCCAGAGCTCAGTCTAGAGGAGGGG  
GCAGCCTTCGGGTTCTCTCCTTCTCCAGAGCCTTCTCCTTTGCACCAG  
GGCAGCCTCTTCTATCTGTTGGAAAGGGCTGTCTGGTCTTGAATATAG  
AGTTCAGGTTTGAGGGGTGTAGGCTGAGGTAAGGCAAACATACATGG  
AATAAAAATTACCCTGTGTCAAGGAACAACCAGAGCTGGACAGTTTTTAA  
ATGTGAAAACCAATTTTATTCAGGACTATGGCGAGAGGTGAAGTAAGACC  
TCAGTATAGAAGTGGGCTCAATTCGAATGCAGCATGGGCAAATGGGAAT  
GTATAGCCTAGGAGCAGGGTGGGAACCTGTGGATGAAGAATTACTAAAG  
GGCATATCAGGGGTGAGGGGGCGTCTGGCTACACCCACTAACTACTGTT  
GCTGAAGAAAGGCTGGTGACATCACTGGGGAATGGTGGGGGATGAAGAA  
TCCAATCAGATGGATATTGAGGATAAGGGGATCTTGATAAACTGGCTTAG  
SAGGGTTTTTGCTAAAACTGGTTTTTCATAGGTAAGTCCACAGACAGGTCT  
TGGAGAAAGTTCAGGGACCTACGGTTTTGTTGCGGCAGATGCTTTGTCATC  
TGTCACACTGGCACTGTCACCTGGCTTTCCTTTAGTCCCTCCCCCCTTT  
TTTTTTTCTGGAGTAGTTTTGGGAGACCAGAGGAGCAGGGAGTTAGGGAG  
AGTAGTCAGAAAAGGCCAGAGAAAATAAGGAGGTGTCTGTAGGGAAAATC  
CTTAAATCCTCTAATTAAATTAATTTAATTTATTTATCTGGGACAAGGTC  
TCACTCTGTTGCCCAGGCTGAAGTGCAGTGGTGTGATCTCGGCTCACTGC  
AGCCTCGACCTCAGGGCTCAAGCAGTTTTGCCACCTCAGCCTCCTGAGTA  
GCTGGGGCTCACAGGTGTGCACTACCATGCCCCGGTAATTTTTGGGTTTT  
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGAGATGAGGTTTCGCCATG  
TTGCCCAGGCTTGGTCTCGAACTCCTAAGTGATCCATCCACGTCGACCTC  
CCAAAGTGCTGAGATTACAGGCATGAGCCACTGTGCCCGGCCTAAATTCT  
CCAATTTTTAAATGCTTCCCTGTTCCCTGTTCCAGATTTGGGATATTGAC  
TGCTGTTAAATCAGCGATTTCTCCCTGTGGAGAGGTAGCCAATAGGAAGC  
AACAAGAGTGAGGAGTCCTTATATCGAAATAGAGGGTAAGAGAAGAGACA  
GATGTTATCTTGGCAGTGATTTAAGAACAGCGAGTCTGTAAAGCAAAGCAA  
AGCAAGGCTCCCAGGTGCTGAGAAACAATGGCTTTCTGGGGAAGCGTCTG  
TGTTTCAGAACCTTAAGTTGGAACATCTCTGAAGATGTTTGCCATGAAGG  
TTTTCTTCTGAAGTTGAGTCTTTCATCACTAGGTAGGCGTGTGTTGGAGT  
CTCTATCAAACAGATCCTGTGTTTATTAGGAAGCTGTGGTTCATAAAGCC  
CCATGCTAATTTTGCAGGTAGCAGGGTGGCCCTGGCCTGACCCGGGGACA

FIG. 3 (36 of 52)

38/118

GAGTGGCTGTCTCTCCCTCCCTCCAGGCAGGAAACTCTCTCTCTGCCACCTAGTCTCTCTG  
CTGCATACCCACATTTTCAAGGGAGCTTCTGGGTGGTGAGTTTACCAGACT  
ATGGTCTGAGGTAGAGTTAAGCAAAACAAAATAAAGTGCATAAAGAAAC  
AGAAAGAAAATCAGGTGTTATAAAAAACAATTTGGCATTGTTTGTGTTTC  
AGCTCCGTGTCTGATTTATTGCTTCCACAAATAGTGCCGATATGCACCAGG  
CACTGTTGTAAAACCTGAAAATATGTTTTTGGATGTGCCAGTCTGTGAGT  
ATTAAACGATGGTTGATTTGAAATTTGCTATGATTCATATTTCTGGGGGT  
AAGATGCAGGATTTCTTTGGGGGGCCTACGATGTGGCATTCTAGAATTCT  
CAAAGAATCAACCCTGGTGGGACCAGGAAGAGCTGAGCTGAGGCCTCTCT  
GCTCATGTGTACTTACTGGAGATCATGGAGACAGGTGAGCCTGAGTGCAC  
GTCTCACCAAAGCCACAGCAGAGGGGGAGGAGGCGAAAGAGAGCTCTCT  
CCATTTCTGAGAAGTTAATGGTAACAATGGCATAACATACCTACTTTACAG  
TTGAAATGGAAACACAGCATTAAAGTGTTTTCCAATGAAATTTGGCAATT  
TGGGAGTTTTCTGAGCTGCATTGGATGTGGTTTTGCATGCTGTTAGGATG  
AGCAAGAGATGATGGAGAACATCTTCTTTTGGAGCTTCTCTTGGACGTG  
GGTCACTCCCACATCATGGAATTAGAAAGCTTAGACCTAGACTTGAATCTC  
ACCTTCTCAAGGTGCTCCCGGGCAAATCACTTAAGATCCATCTTCTCTC  
CTCCTGCTCCTTCTCCTCCTTCTGAGTTTTTTTTTTCTTTCCAAAATTC  
AAATGACACGGTACTGGTAGAAGAAAAGGTCCAAGTCTGCTTTTACAGCT  
CCCCATCCCCAAATGTACTCCGACCCCAAGATGACCATGTTATCATTT  
GATTGACATCCTTCTAGTTTCAACTCATTTCTTGCATGTATATGCACGT  
ACATATACACTATTTTATTTTGGCAGGGGTCACCGTTTAGCTGCATTAAT  
TTCTTATAAAATAATCTATATTTACTTATGGTTTACGTAACAAACATAC  
ACATGTAAGTGTATAGCTTGATAAGTCTTCACTGTAAACCAAAAATAAAA  
TTCGAAGCCCCCCCCAACCGTCTGAATGGACCCCTCTTCTTGGCCAAGAGC  
ATTCGAAAGTTAAACCTGAAAAAAGTTCAGGTGATGGAAGGGAAG  
GTTGGACATGCCCCAGTATACCCCTTCTCCCTTTTGGAAATTCAGGAAAAGC  
TGACCAGCATTAACATCAACACAGACCTTATGTCTGATAGGAACTTTGA  
CAATCTATTCCCTCTGAAGCTTGCTACCCGGAGGCTTCATCTACAAGATA  
AAACCTTGGTCTCCACAACCGCTTATCATAACCCAGACATTCTTTCTGT  
TGAGAATAATTTACCTTGTAACCTGGAAGCTCCCTGCTTCAAGTTCCCTC  
ACCTTTCCAGATTGAACCAATGTAAACCTTACATGCATTGATTGATGTAT  
TATGTCTCCCTAAGATGAATAAAAGCAAGCTGTATGTTGACTGCCTTCAG  
CACAGGTTGTCTAGGACCTCTGAGGCTGGGTACGGATGCATCCTTAACC  
TTGGCAAAATAAACTGTCTAGATTGACTGAGACCTATCTCAGATACTGTT  
GGGTTCAAATATATACTTATGAACTAATACACAAATCAAGTCATAGAA  
TATTTCCATCACTCCTCATCTACCCCAAAATTTCTTATGCGTCTTTGCA  
GTCAACCTCCCAACCCATCCCCAGGCAACTGCAGATCTACTTTTGTCTC  
TGCACCTTCAACTGACCCTTCTGTGATTTTATATGAATGGAATCATGCG  
CTGAGCAGTCTTTTGTGTCTGGCTTCTTTTGTCTCAGCATAATGTTTTTGA  
GGTTTGTCCATGTTTTTGTGTTTGTCAATGGTTAATTTCTCTCCATTGCA  
GAGTAGTTTTCTATTGTACATGTGTACCACAATTTGTATATCCATTCCAT  
TGCTGATGGACATTTGATTTGTTTCCAGATTTTGGCAATTATGAATAGAG  
CTACCATGAACACCCAGGTACAAGTCTTTGTGTGGACTTATGTTTTCATT  
TCTCTTGGAAATGGAAGTGTATATCAATAAGTATATGTTTAACTTTGTAA  
GAAACTGACAACAAATTATCTGCGATGGTTATGCCATTTTGTTTTTCTAC  
CAGCAATACACGAGCATTTTCAAGTGTCTCCACAACCTTTGCCAAAACCTGTT  
TTCTTTAATTTGGACATTTAAGTGGTGTACAGAGGCATCTCATTTGTGGTT  
CTAGTTTTCTTTGCCCTGATGACCAATGGTGTGAACATCTTTTCATGTG  
CTTTTTGACCATTTACATATCCTCTTTTGTGAAGTGTCTGTTCAATATT  
TTTGCCCATTTAAACATTTGGGGGTGTGCTTATTATTGTGTTGGGAGA  
GTTCCATATTTATTTATTTATTTAGATGGAGTCTCACTCTGTTGCCCAGG  
CTAGAGTGCAGTGGCGTGATCTTGGCTCACTGCAACCTCCACTTCTGGG  
TTCAAGCAATTTCTCTGCCTTAGCCTCCTGAGTAGCTGGGATTACAGGCA  
TGTGCCACCACACTGGCTAAGTTTTTGTATTTTGTAGAGATGGGGTTT  
CATCATGTTGGCCAGACTGGTTCGCAATTCCTGACCTCAAGCAATCCACC  
TGCCTCGGCCCTACAAAGTGTCTGGGATTACAAGCATGAGCCACTGTGCCT  
GGCCCATATTTATTTTTTATTCTTTATTTTGTATACAAGTTCTTGGTCAG  
ATACAATAATACCTGGTCAGATGAGATAATGAGTTGGAAAATGCTTTGCA  
AATGGGGGAGAATAATTTAAATGTTATTTATTTATTAAGAGCAGAGGCCC

FIG. 3 (37 of 52)

TTCCCTGTTGCGGTAC...AAGCCGTTTGCTTCTTCTGCCTTTTATAAA...  
AGCAGAGTCGAGCTACACAGGCTGTCTGTGTGGCTGCTATTAGTTAATC  
AGAGAGTTTTTTTTTTTCTTGCCTTGTCTTAATTTGTGACACATAATT  
AGCCACAATATGTGTTTTTCAGTTGTGACACTGGCCTGGGAAACCAAGGGA  
TGTTTAGAGTGGATTTCCTTGATTTTGCAATAATTGTGTGTTTTCTGCA  
TCTTCTGTTAAACACAAATTCATGGAAGCAAAACATGGAAGCAAAGTACC  
CTGGACATCCCCCTTCTTTATGAAATTGATTTCTCTTAAATGTAATGTT  
TGCTTGTTCCCTTACTTTAAAAGCAATTTAAGAGTTTATTGAGAAAGTGA  
GCCCTGGAAACATAGATGCATAGAGAGAAAATTCTACCACCTCAGGTCC  
CTATTGTCTTCTCTCATAAAGTGTAGTTTCAGGGCCTTTTAGAAGTTTCT  
TTTCTGCTCTGATTGTCATGTTTGTGAGTGTGCTATTTTAAGTATTTGG  
ATTTGGTCTGCAAAATCCTATGAGAGATGGCAACAGAGTAGGGATCTCAAA  
GCCTGCAGGTTGTATTAAGTCCAGCAGGGCCTTGTATTTACAACAGAGGG  
TCCTTGAAGACATTCCATATATTATGCTAGGGGAGTGGCCAAGCAAACCT  
TAATGTGTCCCTATGGTGGGATATTTGGGGTTAATACCTGCCCTTCTCTT  
AATTTCTTTTTCTTTTCTTTTTCTTTTTCTTTCTTTTTTTTTTTTGA  
TGTAAGTCTTGCTTTGTACCCANGCTGGATTGGAGTGCAGTGGTATGATC  
TCAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTC  
AGCCTCCCAAGTAGCTGGGACTATAGGCACACACCACCATGCCTGGCTAG  
TTTTTTTTTTTTTTTTTGAACNGAATCTCGCTCTGTGCCCCAGGCGGGA  
CTGCGGACTGCAGTGGCGCAATCTCGG

>Cont:ig39

CGCTCGCATCCCTCATATCCATGAGTGTCTGTGGGGCCTGCCTCTGAAA  
TAAATCCTGCCTTTGTCTCCAGTTCACCTCCAGCCACCCATCCTGGGGCT  
GCACCTCCTCCTTCCAAGCCCTCTCCCTTTCTTCTGCTGCTGCCTGT  
CATGTCAAGCATATGCATCAGTGCAGACCAGGACATTTGAAATGCAACCAG  
TACAATTGGGCGCGGTTATGCCTACCAGTTTTTCTTCTTAAACATTTTA  
TATTTATGTTTGAAAGCATGCCACCTTCTTCACTTGCCAACCTTGACAGA  
TTTATTAGTTGACAACATCCGCTGATAGCATCAGTAATAAGTTAATTGTT  
TTTGACATGTAGCTTTAATTATTCTCATTATCATTATAGGAGTTATTC  
TTTGTAAGGGTAACTGAGTTTCCAAAACAAACAGAAATTGGGGTGGG  
CCCATGGAGCGTGACTCATGAAATCAGATTCTTAGAAGGACCTCGGCAAG  
TCTCTGGGTTGCTGTTAATGAGCCTGGCTGGCTGCCAGGGGTGTGTCTGC  
CCTTTATGAGGCCACCACTGTTCAAATGCTTGCTGCAGCATTACTTGCC  
TAGGTAGTGCTTGTCTACTGAACTGTCAGGGATCCAATTCTTTGTGGT  
CTAAGTAACAATACTCAGATTCACAAGGAATTGATTAATAAGCCAGAATG  
CCAATGTATTACATTTTTGATGAAGACCATATTTACAGTGATTGTATCTG  
CTCAAGCTCAAATTAGGATTAGAGTTCTGACAAATACATATGTGAGAAGT  
ATGAGGTTAAATACTTGAAATTGGACTTTTCTAGAAAATCTGAATGTGA  
TTGCCATTACATACCTTTCTGGGGATGATGATTCTGTACTTTTATTTT  
AAAAGACATAGAAAACCTAAGTAAGTACAGATTGCTTGGCTGGGCACAG  
TGGCTCATGCCTGTAATGCCAGCACTTTGGGAGGCCAAGGTGAGTGGATT  
GCTTGAGCTCAGGAGTTTGAGATCAGCCTGGGCAACATGGTGAAATCCCA  
TCTCTACCAAAAATACAAAAAACAACCAAAAAGAATAAA  
TTAGCTAGGTGTGATGGTGCCTGCTTGTAGTTCCAGCTACTTGGGAGGAT  
GAGGTGGAAGAATTGCTTGAGCCCAGGAGGTGGAGGTTTCAGTGAGCTGG  
GGTTGCAACAGTGTACTCCAGCCTGGGCGATAGAGTGAGACTCCGTCTCA  
AAAAAATAAATCAGATTGCTTTATTGCTGGTCTTTCTTTCTAAACTGA  
GATTGGGTCCCATCATCCCTGGCCCCCATTGGTTAATGGTTCTCCTTT  
GTCTATTGAATAAAATACAGATGTCTGCTTTTGGCAACATGGTTGAATGT  
AGACACTGCAGGGTCTTCTGACTCAAATGAGTAAGGCTTAGATAAAAC  
ACATTTTGAAATGCATTTCTGGATGAACAGCAAGGAAAGGAGATCTCTTA  
AAATCCTCTTTCTGTTCCCTCTCCCTACCCCTCCAAGTGGGCTTAAGT  
AGGAAGGGTGGTGAGCGGCAGGTAAACACACGTCAAAGGCAGTCTTCCTC  
TCTGAGGGAAAACACTTGTATAAGCATTGCAATCAATGGGCCTCTTTAAT  
TATGTGCCAGTGGCAAGAGCGGGTGCTGAACCCAGGGGCCTGCCTCAATC  
CGGGGCCTTTGAGGCAGAATAAAGTGGTCTCAGGTTGTTGGCATTTCCTT  
GCCCTTCCACCCGAAGCAGACACAAATCCTCTCTGGAGGCAAGTTCCCA  
ATTGAGCCAGTACAACCTCCACAGACTAAGATCAATCATGTACAAGCTCA  
CAGACAAAGGTCAACCAACACACAGAGCAATAAACAAATTCATGAGTGAC

GTGAATGAGAATAAACACAAACAATAACCACCAGCTGGGATGCTCTAAG  
CTTCAGCTGTTAGAATTCTTGAATATAGAATAAACTGCCACAATGGCAA  
ACATGCATCTAGTACTTACTGTGTGCTGGGTTCTAAGAATTTTGCACATT  
GTGCCAGATACCGACTCAGCTTCACACTCACCTCCTACTGTGCCCTCTT  
AATTTGCACTAGATTAAAGGTAGAAAGGAAGAGGCAGCTATTCTGTTCT  
TGGCTGTGCTCTGGCAGCACATGCAAAATGGGCAGTAACAGTGGCAGTC  
ACAGGTAAGTAGCCTTCTCACAGTGTGGAGTTAAAGGCATGGGACTGAGA  
CGAGCAAGGTTCTTAAAGGGACAGTGGCCAGTAGATGACCAGGGGCTACT  
GGAGTGGCTGCATGGCTCTGTGGAAGCTCAGAGGAGCCTTGGGTCTTGA  
GGTGCAGTAGCAGCTTTCTGTAGTTCTTGATCTCTGGGTCCCAATCTT  
CCCCGTTTTTGCTCCTCCACTTCTAATTTTGTAACTGACTTCCCTGTGTG  
TACTTCTCTCTGATTGAAATAGCCAGACTGGTTTCTGTTTCTGATAA  
GACATTGTCTGTACGAACACAGTAACCTCATTTAATCCGATATCTCTATG  
AAGGAGGTACAATAATTATTCCTATTTTACAGATGAGGAAACACAGCAGA  
GAAATAAAGTCAATTGTCTAAGGTTGCACATTTAGTCAAGGGAAGGTTG  
ATATAACATATAATTATTTAGAAAACATCTAAGGAAATAAAAGGCATAAT  
TTAAAAATAAACTAGGCAGGTTTAAAAAAATGAAGTAATCTATAAGTAA  
AAAAGTATAATTGTTGAAATACATATCTTAGTGGATGGGTTAAATAGCTG  
AAGAAATGATTAATGAAGTGAAGGTAGTTCTGAGGAAATCAGAATTCAG  
CATAGATAGAAAAATGGGAATTTACAAAAGTACACAGGAATTATAAAAG  
AGGTTAAATTATAGGGAGGGTAGAATGAGAATTAACATTGGTCTAACTGG  
AATTTTGAAGAAGAGAATAGAGAGAATGAACAAGGCAATATTTAAAGAG  
GTGGCTGAGAATTTTTCAGAACCAACACAACTATGACTTTACCAGTAGA  
GAAAACAATGTACACTGAGGAGGATAAAATAAATATACTATGAACAAATTG  
TAATAATAACTCAACAAAGACAAAGAGAAGATGTTAAATCAGCAAAA  
AAAGAAAGTCAGACTTAGAAAGAAATGACAATGGCAGACTACTCAACAAC  
AACAATGGAATCCAAATTCGGTCAAACAGTATTTCTTCATGCTAGCATA  
TAGC

>Contig40

GGGAGTCCGCTATGCTCCTAAAGATTTGCACCTCTGATCTGGTTTGTAGT  
TAGTCTCTTTTATTGCTTTATCCTACTCAACTAATTTTTTTAGTGCCTGT  
TTTTTTTTTTTAAATGTGTGTTGATGACTACAATTCTAAACTCATTCTA  
CTGATTCATGGGTGCTTTAAATCTGAGCAGTCTTTCGCATTTACTGCCT  
GTGATGGCCCATCCCACCAGCTAAAGTGTGTGGCCACTGCTTACAGCACC  
ATGTGATAACGAGTAAGGGAGAGATGCCGCCAGACTCTTCTAGGAGCAG  
CCAGTAGGACCTTCCAGGGGTTGCAAGCAAACCACAGCAATATGTGGAGT  
GTGGCAGAGGATGGCCCAAGAGGATGTGGCAGCGGCTAGTGCAGCTCAG  
CTTAGTCTGAGAGGAAATGCTGGAGAGGAGAGCCAGTCTGTACAGGCAT  
GACAGCCACAAGGACTTCAACAGCTAACATGGCTGAGTGGACTTTATGTG  
CTATCTCATTGAGAAAACAGGAGCAATCAGAAAGGAGTCACCTCCTATTT  
GTACCCAGGAATTGCTAACCTACTTGCACTCTGAATGATGTCCATCACTT  
CCCTTCATCACCTCCTCTGGGGGCTCTGCAAGGATTTGACTCCTGCATTA  
GTGATCTGTCTCACCTACGTTGTGATTACATGAACCTTACTAATGTGCTA  
TGTGACAACCTACCATCTTAAACACAAAAACCTCTTTTGATTCTGTGGCT  
CCCTCCAGCTACCCCTGCATTTCTCTGTCCCCCTGCCCGTCTCTGCACT  
CACTTTTATTTTACAGCAAACTACTCAAGGGAGTCTCAGTGCTCCTTGG  
CTCCATGTCTCCACCTTTCATTCTCTCCTCAGTTCACTCCTGTCAGGCTT  
CCGTCCTCAAGCTCTTCTTCACTTTTGTCTAGGGCCGCTGACATCCTCT  
TTCTTGCCAAATTCAGTGGCCAGGTCCTCACTTACTCAACTGCTCAGCAT  
TGTTGGGCTGGTGGACCACATTCTCCTTCAACCCTTTTGCTGCTCTC  
TCTTCTCTCAGATGTTTCTCTCTCTCTCACTGGCTACTCCTCTTTTGTCT  
CCTTTGTTAGCTCCATTTCTTCTTCCAACCTCACTGTGCTGGTGTGCCC  
AGTGCTCAGTTTTTAGCTATTCTCTCTTTTCCAGTGGCATTATTAGATG  
GTATCATGTGACCCATGGCATTATATGCCTTCTACATGACAGTTACTCCT  
GAATATGAATCTCAGGAAAGATTGGATTTATTTTAAATTAATTTTTTTA  
AATTTTATTTTAAATAAATGAGGTCTCTCTCTGTCTATCCAGGCTGGAGTGT  
AGTATTGAGTGATGTGATTATAGCTCACTGCAGCCTTGAACCATGGGCTC  
AAGTGATCCTCCTGCCTCAGCTTCTGAGTAGCTGGGACTACAGGCATGT  
GCCACCATGCCTGGATGACTTTTGTGTGTGTGTGTGTGTGGAGACAG  
GGTCTTGTCTTATTGCCAGGCTGATCACAACCTCCTGGCCTCAAGTGAT

CCTCTCACCTCAGCCT...CAAAGTGCTGGGATTACAGGTGTGAGACCA...  
CTGGGCTAAGATTTCAGATTTTGTATTCAATTGACTGTTTGACATCTTCAC  
TTGGACACCTAAGAGGTATCTCAAATATTAATTAACCTGGCCAAAATACA  
GAACCTTTTGACCCCTGCCCCACAATACTTGCCCTTCCCCAGACTTCTC  
CATTTCTGTTAAATATCCCCAGTTACTCAACCCTCAAACCTATGAATGCC  
CTTTGATTTCTTTCTTTCCCTCATCTCCTACGTTGACGCCATCAGCTAGT  
TTTGTGTCCTTTATGCCCAGAATATAATCCTCACCACCTTCTCTCCTATT  
GCCCGAGTATAAGATGTCAGTTTTTCTGTCACAGTCCATTGCCCTGACCT  
CCTGAGTGGTTTGCTTCCACTTTTGACATTTGTATTCTCTTTCCCCCAG  
GGTCAATTTTTCACAGCAAGAGTGGCATTFTTTTTTTTTTTTTTTTTTTG  
AGACGGAGTCTCGCTCTGTGCCCCAGGCCGACTGCGGACTGCAGTGGCG  
CAATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACGCCATTCTCCT  
GCCTCAGCCTCCCGAGTAGCTGGGAATACAGGCGCCCCGCCACCGCGCCCC  
GCTAATTTTGTATTTTTAGTAGAGACGGGGTTTACCTTGTAGCCAG  
GATGGTCTCGATCTCCTGACCTCATGATCCACCCGCTCGGCCTCCCAA  
GTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCAAGAGTGGCATT  
TTAAAACCATATATTAGATCATTGCTTTTGTGTTTGGGAACCTCCAAGGG  
CTTTGCATCATATATCAAGTTGACACCTCTCCTACCCAAGCCTGGCTCTT  
TCCTGCTCCTCTGTCCTCTCAGCCCCCTCCACCCATTGTTTATGCTGCTTC  
AGCCACACTGGCCTTCTTGCCATGCCACATTTGTGCTAAGCCACATCCA  
ATCTCGGGGCCCTTGCACTCGCATTTCTCTGCTTGGCATGCTGTACCCC  
AGATCTTTTCATGATTGGCAGCTTCTGTACATTACGCCACCTGCTCAAGCC  
ACCTTTTCAGAGGGCCTTCCCTGGCCACCTCACCTGAAATAGCACCTCCG  
ATTGCACCCATCCGGTTATTCTCCATCCTGTTCTCTTGGCTTGGTGATTT  
CCATCACTGATGAGGAAATGAACCATGGAATGCTAGGGCTGATGACCAGA  
ACTTTCCCCCACCACCATATTACAGAGGAGGAAATGAGGTGCGAGGT  
AAGATGGGCCCAGGATTTCTACTCCCGCTGGACTGCAGGCACAGCACTG  
ACCTCAGCTGTGCTCACTCTTGGCATTACCCAACCTTCTATCTCCAAC  
TGCCCCATTTACCAGAAAGTGAAATGTTCTCAGAGACGGTGAGCCACCTG  
ACTTGACAGCAGCCCAGGGCCCCCTGGCACCTGCTTTCTTCTCCTGC  
CATCCTTTCTCTCCAAGACCTACCTTTCCCTGTGATTCTTGCCACATG  
CTGCATTTCTGTTTATGACCTGATTTCTGAGAGGGATTTGAATTTTC  
ATGATTATTTATGTAAGCAAATCATTATGCTTATACAAATGAGAAAAGGA  
GTGCTTCTGGACTTCCAGGGACAAAATCTTGTCATTGGCTTGCTTTCA  
TATTGCTAATTAAGGACCCAGGATGTGGGTGAGATGTGCTAAAAGCTGAG  
AGGAGGCTCTGGACTCTGACTATGGGCCCACACCCCTGGGCAGGCATCAC  
ACTAGTCTTTAGGTATCTCAACCCAGCTTCCAGTTGAATCAGATGTT  
TGTGAATAACTCAGCAAGGCTGTATGGGAAATGAAGAATGAGGTGGGGAA  
GAGGCCTGTGAGAAGACACACTGACTTACCCCTCTACCTCTAACTAGGG  
TGTTGTAGCAGCCACCCACCCACCAAGTCTGTCTTCCAGACCACGTATGC  
TTTCTTCCACCTTTGTCATCTTTTATCTTCTGCCAGCCAGATGCTTGCTG  
ACTCCAGCCCAAGCCTATAGGATAAGCTACAGCCTGTCCCTACAGACTAC  
GCATTGCAGAATCTAAGACATCAAGTCAAGTTCGGAAGCACTTGCTTCT  
CCTCTCCAGGTACACAGGCTCTCCTGGAAAGCTGGTAGCAGCTGTGGAGG  
TGTGGTGTGTTACCTGCTGCAGGTGCAGAGAAGTTGACTTCACAGCCCTT  
CAGAAAGACTGCCTTCTTCCAGTTGTATTTGTGTACTTGCTTGGGTGTG  
GGAGGATTCTCAGCTTTCTCCACTCAAATTATCAGACCCTTTCCATTTAG  
TGGTAGACCATTTCCCTCGTCCAGGCCAAGGGCACATAGTACAGAGAAAT  
AGGGAGTTGTTACCCAGGGAGAGAACTTGCTCTAAACCTGTAATAGAAA  
GGTCAGTTCTGGTCTGGAGGGTCAATTTTGATCTTTGGCTCAGATCCAGG  
AATTGGAACCAAGGCTTTTGAACATTTTAATGCAGGGGATTAAAAAATG  
ATACGAGTCATTACGAATATATTTGCTTAACATCTAAAGAGATCCCTCA  
AAACACTAGAAAAAATAAGAACAAAAATCTAATAAAACAAAATTTGTAA  
ACACATTTACCAAATTTTTTTTTTTTGGTAAAAATTCAAATGTCATAAATA  
AAGCTAAAGTTCTCTTGATGACTCGCTCCTCTGCCCTATTCCACTCCAA  
GTAACCACTATTATCAGTCTTGCCAATACCCTTCCAGACCTCTCTACCTC  
TATATACCATTAGAAGCACATGGTTTTGTCATTGAGGATGTGCAGTGT  
GTTTTACGTAAATGTTATCACTCTGTTCTTGTTCATAATTTGCCTTTTT  
CTCTCAATGATTTGCTTGGCTATCTTTCTATTTTCACTAGCATCTCCTTC  
TTTTTAACCTTACCATTTGTTTATTTAACCTTGCCTCTATCAACAGATATGT

FIG. 3 (40 of 52)

42/118

AGGTTGTTTTCTAGTTGA.TTCATTAAGTATTTATAAACCAACGCATCAG1A  
BATGTCCATAAATTTCTTTACGGAAGATGGCAAGTAGTGGAATTGCTGAG  
CCAAAGAACATGTTTTAAAAAACC2AAAAAACTAGACGCTACCAATTTTC  
TCTCCAAAATGGCCATACCCACTTACCCATACAGAGATGATTTGGAATCT  
GGCTTCTCACAAGGTGAGATGCCTTCACAGTTTCATTCTTCCTGGCATG  
TCTTCCCTTTTGTATCTGAGAGAGCTGGCAGAATTGTGTCACTAAATCAA  
GGATAGAGGGTCAAATGACAGCTCAAGCTCACAGGCACCTCTGCTTTCTT  
CCCAGACCACCTGCTTTCTGCCACCAGCTCTGTTCCATCTTATAGAATG  
GTTGCCACTTGGGTGTCTGCTCCGACAGCCATGTTCATCCTTTGCACTGCA  
GTTATGAAGCAGACAGAGCTAGGAGAGGGGCTTTGCCAGCCTCTGCCCTA  
GCTTGGAGAATTTCAAAGAAGGAGGGTATTGAGAGTGAGCTGCCGAAGAC  
TGGCAGCTCCCTCAACTCAACAGTTGTCTTCCACAAGAAGTCAGATACA  
TTTTTTTGGGATAAAATATTTAAAAATTATTATTTTATTTCTGAATAATA  
TATTTACATGATTCAAAAATCAAACTGTAGGCCAGGCATGGCTGCTTATG  
CCTGTAATCCTAGCAATTTAGGAGGCCGAGGCGGGAGGATCACTTCAGCC  
CAGGAGTTCAAGACCAGCCTGGGTAACATAGTGAGACCCTGTATCTACAA  
AAATTTAAAAACAAAATTAGTTGGGCATGGTGGCTGATATGGTTTGGCT  
CTGTGACCCAACTCAAACCTCATGTTGAATTTTAATCCTCAATGTTGAGG  
GAGGGTCTGTTGGGAGGTGATTGGATCATGGGGGTGGGTTCTCCCTTGC  
TGTTCTCATGATAGTGAGTGAGTTCTCACAAGACCTGGTTATTTGAAAGT  
GTGTAGCACCTCCCCCTTCACTCTCTCACTCTCCTGCTCCGCCATAGTAA  
GATGTGTGTGTTTTCCCTTTGCTTCCGCCATGATTGTAAGTTTCTTGAA  
GCCTCCCAGCTATGCTTCTGTACAGCCTGTAGAAGCTGTGAATCAGTTAG  
ACCTCTTTTCTTCATAAATTACCCAGTCTCAGGTCAATCTTTTATAGCAGT  
GTGAGAGTGGATGAATATAGTGCCATATGTTTGTATTCCCAGCTACCCAG  
GAGGCTGAGGTAAGAGGATTGCTTGAGCCTGGGAGTTAAGGCTGCAGTG  
AGCCATGACTGTACCCTGCTCTCCAGCCTGGGTGACAGCGAGACCTTGT  
CTCCAAAAAATAAACCCTGTGTAAAATGTGTTTCAATAAAGTGTC  
TTGCTCCACACCTGTCCCTATATATCTTATTCCTCAGCCTCCGACAACCT  
ACTTTATTCATTTCTTATGTATCTTCCAGAATCAAAAAAAAAAATCAAA  
TACAAGCACAGTGGAAATGTATTGCCCTTCTTCCCTCCCTTTTGTACAT  
CAGAGTTAGCATATCATAAATACGGTCTGCATTTTCTTCTTTTTCAGCTA  
TCAGCATGTTTGGAGAGGATTTCAATTCGTGCAGACAGCATGTATTAG  
TCAGTCTTTGCATTGCTATAAGGAAATACCTGAGACTGCATAATTTATAA  
AGAAAAGAGGTTTAAATTGGCTCACAGCTTCGAGGCTGTTCCACAGGAAG  
CATGGCAGCATCTGCTTCTGGGGAGGCCTTAGGAAGCTTTTACTCATGCA  
GAAGACAAAGCGGGAGTGATGTCTTATATGGCAGGAGCAGGACTGAGAG  
AGAGAGAGAGAGAGAGAAAGGATGCCACATACTTTTAAACAACCAGATCT  
TGTGGGAAGCTCTGTACGAGAACAGCACCAAAGGGATAGTGCTAAACCAT  
TCATAAGAACTCCACCCCATGATCCAATCACCCACACCAGGCCCCACC  
TCCAACATCGGGGATTACAATTTGACATGAGATTTGGGCTGGGACACAGA  
ACCAAACAATACCAGAGTGCTTTCTCATTCTTTTCTATAGCTGCCTAGTA  
TTCTATGTCTTTTACTTCAATTTAGGCAGTCTCTTGTGATAGACACTTGG  
GTTACTTCCAATTTTTCTATTACAAATGATGTGCAATGAATAATTTTGA  
TCATTTTCCATTTACATGGGTTATGTCCATCTGTGGGATAAATCTCCAG  
GAGTGAAATTGCTGGATCAAAGGGGAAGTGCACTTGTGATTTTTCATAGTT  
AGCAAATTTTGTCTATAAGGGTCATATCAATTTATAGTCCCACGCGTAA  
TATTTAACAGTGGGGATTTCCCGACAGTTTGACCAACAAGGTCTGTTGTT  
AAACTTTTGAATTTTGTCAATCTGATGGGAAAATACTAGTATCTCAAAGT  
GCTTTTAATTTGACTTTCTTATTACAATGTTAAGCATCATTTTACTCTGC  
CCAAGATCAAATAGTATTTTCTTTTCTGTGAACAGACTGTTAAGATCCCT  
TGCTCTTTTGTGCTGGATTTTGTCTTTTTTTTCAAATGTTTTGAGG  
CAGTCTTTTACATGTGAACAAGTTATCTCTTTATCTGGGGTGTGAGTTA  
CAACTACTTTTCTCTGGCTTGTGTTGCGCTTTGACTTTGCTTCTGGTGA  
TTCCCGCAATTCTGAAAGTGTAATTTTGCATCATTCATTCTTATACACC  
CATGCTCTTGTTCACGCTGGTTCTCTACCTGAGGGCTTTTTCTTTTCTG  
CTTCTATCTGGGAACATTTTTTTGAGAGAGAGTCTCACTCTCTCGCCCAG  
GCTGGAGTAGTGCAATGGCGCGATCTTAGCTCACTGCAACCTCCACCTCC  
TGGGTTCAAGCAATTTCTCTGCTCAGCCTCCCAAGTAGCTGGGATTACA  
GGAGCCCCACCAAGCCAGCTAATTTGTTGATTTATTTATTTATTTT

FIG. 3 (41 of 52)

43/118

TGTAGAGATGGGAGTC. . . ACTATGTTGCCAGGCTGGTCTTGAACTCC. . .  
GGCTCAAGCGATCCACCCACCTCGGCCACCCAAAGTGCTGGGATTACAGG  
CGTAAGCCACCATGCCCAGCCCATGTGTGGAAATCTTCTGTTTATCCCTT  
TAGGCTTGATTCTTATGTCTGTTCTCCTCCCTCCTTCTGATACTCCTCT  
TGTTCTTTATCTTACTCTACTTGTCTGTTACCTTGTTTCTGCTTATAAC  
TAGCTGCCCTCTCCTATCTGAGGAGGGACTTGTGACTGTTCTCATCTCTGT  
ACTCCAGCTCCTAGTACATAGCGCTTGCTCAACAGATGTTTGGTGCATT  
GATAGATAAATCACTGGTAGCTGTTACTACCAGTCCTGACTCCCTGCAGT  
GCTTCAGCTGATCCTGTTCCAGATGTGCACTGAATATCCTTCTGTTGAAC  
AACAGAAATAAAGGGGATGGGTGAGGAGGATAGTCTTCGGTGGCCAAGGA  
TATTTTATAGGTACTTTGCAGCACTCAGCAATGAGGAGTGGGCTTTAGTCC  
CCCAAGAACTCTCACAGCCCTGGGTGTCTTTACTGTTCACTGTCAAATCC  
AAGACAAGTCAATGATCAGGAAAGACCATTTTTTTTTTGTTCAGTGAAAGTT  
TATTTCAAGATCATTGAACAGTATGATATTTGGTAATTTTATAAATATTC  
CCACTTAAATGATCGGAGCAGATATATTTTCACTCGTAATTAAGGACA  
TGATTTAAAGAGAGCACACCAGTCCAAATTGAAATGATTCCATAGCTATT  
AAAAAACTAGGGTTTTTTTACAGACAATGATACTTTTGGCCCCCTTTGAAT  
AGATTAGACCAATGAATAAAACAAACAAACAAATAAATAAATAAATAGGG  
AAGCGGTTGCTCATCAGAATGTGGGAGCGAATGACAGAGGGTTTCTTAGA  
ACCAAATGTGGCCGTGGTTTTCTGTGAGCGTGCTTTAAGTGAGTAGGAGA  
GGTGAGAGAGGCGCTGGCTCAACAAAAGGGCTGGGGATTGTCCCTGAAGAA  
CCAGAGCTGANTTNCATCAGGAGTAACANAGGTAGATAG

>Contig41

CCGCGTTGAGGTTCCACGCAGTTCAAATTATGTCCAATTATCAACATTAA  
TGCACATTTTCAATAGAACCTGTTCCGGCTTTTCTTAGGAGGGGGCGGG  
GAGACGTTGTTCTCTGGGAATAAGTGACGCAGGAGGCTGAGAAGGCTTC  
ATTCCATAGCATTCACTTACCTCCAGCTGTAGAGTGGGCTTATCATCTTT  
CAACACGCAGGACAGGTACAGATTTTTTTCTTTGAGGCCCAAGGCCACAG  
GTATTTTGTCTTACTTTCTTCTCCTTGTAACAAAGGACATGGAGAACACC  
ACTGAAGAAAGAAGGGGGTCTTGTGGTTAGGGACACAGCAGTGCAGGGTC  
ACCCCAACCCCTAGGCCCATGAGTAGGATACATGTAATTTGGTAGCCTC  
TGTGGGAACCCACAGTGAGGTTTCTTGGCCTAAGACACAGGATAACTTGA  
CTTCTCACAGACAATAGCAGGGTCATTTTGTGATTTAGGGTTTCCCTC  
AAAGGCCTGAGGGTTTCTCAGAGCCTCATAGCAGTAGGAACGGAGAATGA  
AAGAGGGTCTACATTTTAAATGCTGAAGGAAGGAAGGAAGGAAGCCATTG  
TGCTACTGGCTGGCAATGTGCCATCCACAGGAGCGGAACAACTTGATCA  
ATGTGGAAGGAAAGGAAAGAGGTGAGGCTGTACTTCTGCCAGAAATCAGG  
CACCAGAACTGTTTCAGGAACAGAGAGTAGCCCATGGGAAGAACTGGGA  
GAGGAGAGGCTGAGCTGGGAAAGTGGCTCCAAAGAGAGACACTCATTTTG  
ATCTTCTCAGTCACAGCAGTGTCAATTGGAGGCCCTGGGATCACTCTTA  
CTACCCGATTCCAAAGAAACAGGATTTTCTTGGCCTGGCTGAGAGCAAAT  
AGCTTCCCCCTGAGTGAGGCTGTCTTCAAAGTCAGCAGCCTTAGTTGCC  
CACACTCTGTGTCAGAGGCTTTGGCTACTGTGGCAGATGCCAGGCAGAT  
CACCACAGCTAATGATGGGTTTACCAGCACTTGAACTTTTGGCCGTTACA  
GCGGAGAGATATAAGTTCTGTGCTGGGCGGTAAAATTTCCCTACAAGGAAC  
CACCTGGCATTGGGTGGGACGGATGTTGGGGCAAGGGGGGAAGACTGGGG  
AGGGGGATGGACACATTATCGCTCCAGCACTCTTGTTTCAGCCTCAACAA  
CAGGAAGAGAGAACCCACAGGCAGTTAGGCCATGTCCATCAAATGACCCC  
ATATTGTGGAAGAAATTGACATTGCACTATGCCCAAGAGACTTGGGTGGAC  
ATGGTCCTGGGAGTGCTTGAGCCGTCTAATTTCTCAGGGTCACACTCCTG  
TTAACAATATGCACTGGCCAGTGCAATCAAATGTGCCATTTCTAGGACCAA  
AGTTTGAATATCTTTTAAATATTTTCTTCACTTGTGTTGATCATTTG  
CCTTAAATTAACCTTTCTACTTTTGTAAACATGGAGAATTAGCAAGCTG  
CCAGGAGGCCAGGCAGGGAAACCAGGATGTTTCCATTTACCTTGTGCTC  
CATATCCTGTCCCTGGAGGTGGAGAGCTTTCAGTTCATATGGACCAGACA  
TCACCAAGCTTTTTTGTGTGAGTCCCGAGCGTGCACTTCAGTGATCGT  
ACAGGTGCATCGTGACATAAGCTTCGTTATCCCATGTGTGCAAGAAGAT  
AGGTTCTGAAATGTGGAGCACATGTTGTTTAGGTATAAAATCAGAAGGGC  
AGGCCTCTGAGGCGAGGTGGCAAAATTTGATTTCTTGGAGGACACCTGA  
GCATATACGGTCAAAGTCTGATGACAACACCAGTAGGGATGAAGCTGGGA

GTGGGGTGGCTAAGAAC CTGGACCTGACACTATTAGACATGGGTTCC  
CTTCAGGTCTATTACTGCTCACTGTGGCCGAGCAACAGAGCTACTTAGGT  
AAAAATGGTGATGGTCATAACACTAGCCACAGGGAGGTTACGAACCTCTG  
GTGACCAATGTAAGTGAAAGGCCCTGAGAAAGAGTGAGGGAGTTGCAAAT  
GTCAGTAGCCATCAAGATCTTCTTTAAGAATAGTTTCCACTAAAGAGATG  
ATTGCTTTGGTTTCCAGCCTTCTTTGTTTGTCTCCCCGCTGGGCCTTCT  
ACCTTTAAAGGGCTTTGGCTCTGGGGGAATTGAGTTGGCTGGGGCTTGAT  
GACTTCCAAGAGGACACAAGTGGAGATCTACTGCCTGCTCTTGGCTAACT  
ACCTTCTTCAAAGATGAAGGGAAAGAAGGTGCTCAGGTCATTCTCTGGA  
AGGTCTGTGGGCAGGGAACCAGCATCTTCCTCAGCTTGTCCATGGCCACA  
ACAACTGACGCGGCCTGCCTGAAGCCCTTGTGTAGTGGTGGTGGGAGAT  
TCGTAGCTGGATGCCGCCATCCAGAGGGCAGAGGTCCAGGTCTTGGGAAGG  
AGCACTGCGGAGAGAGCGAGGGAGGGAGCCTGGTGAGGTGGTCTTGCCAG  
GAACCATGCTTTGACATCAGAGAGTAGAAAGCTCAGAGAGGAGGAAAGGG  
CTTGAAAGAATCCCGAGCTTCTAAAGATCATCCCTCTCTGGGCCAGGCGT  
GGTGGCTCATGCCTGTAATCCAGCACTTTGGGAAGCCGAGGTGGATGAA  
TCATTTAGGTGAGGACTTCAAAACCAGCCTGGCCAACATGGCGAAACCCC  
TTCTCTACTAAAAATACAAAATTAGCTGGGTGTGGTGGGGTGACCTGT  
AATCCTAGCTATTGAGGAGACTGAGGAAGGAGAATCGCTTGAAGTCAAGG  
GGTGGAGGATGCAGTAAGCCAAGATTGTACCACTGCACTCCAGCCTGGGC  
AACAGAGTGAGACTCTGTCTATAAAACAAAACAAAACAAAACAA  
AATAAAATAAAATAAAATAAAAGATTATCCCTCTCTGAAGCTCAAGGAG  
GTTAAGGGTGACTCAAGGGCACACAGCAGGTTAGAGGCAGACTCAAGAT  
TAGAATGTGGGCTTTCTGACACCTTACAGGCTATTCTTTTGAATAAATC  
CCATTTCTACTTTGTTCATCTTTTTTGTACATGCCCCACCTACACCATAC  
ATGTATACCTTCTCTATATCTTTTTGTATCCCTAATGCTGTCACTATG  
ATTTGCTTTTTTTCATGCAGATGACCATAACATTTTCCATTACCTATGCTC  
ACTCAGCAAGTATTCAATTTTTTCTACACTGTTCTTTTTTTTCTTTTTCA  
TAACACTGTCTCATAGGCATTCTGCAAATCCTGTGAGAGTACTTTTTGTG  
AAATGTTACCACTTTCTCTTATTGAGAGAAGCTCCGTATTAAGGCTTCA  
CTGAGGTTGCCTTAAGGCATGATAATGGTTCAAAGGCTTGAAAGACAGTT  
AAAGAGACCTGTAAGTGACAAAAGAAAGTTGAGCAGGAGAGAATTTCT  
GCCTGGAGCAGCAAGCTGCTGGAAGAGGCAATGGGGGCAAGGCCAG  
GCAGACAAGCCAATGGGCTCCTCCACAGCTGCAGCCAACAAGTTATGCC  
AGTCTTAAACTTCTAAAGAAATATGTTTTTAACAAGATTGAGGACTGGA  
TTATGAGGCTAGGGGAGGCTATCACAACTGGAATAAAATAAAGCCAGAG  
AAAAGTGGCTGCCTTCCAACCTGCACAACTGACCTAGCTAGGCTGATGGC  
TGGGCCACCTAGGAAGGCTACTGAGCATCATATAAAACAGAAGGGACAGC  
AGGAATATAACATGGCTCTTTGTAAGGATGAGTCTGAAAAATGACCATT  
GCTGCCCAAATGCCCTTAGCTACAACCTGAAAATATTTTCAAGACTGGAGGT  
TGCAGGATGCTGGAATCTCAGAGATCATCCAGCTCAGCCCTTTATTTTTT  
AGATGAGGTCCAAGCGGGTAAATGACTTGTCAAGGTCAAACAGCAAGT  
GAATGGTTTTCTTTCAAGTCTCAATTCATCTTTTTGTTTATATCATCTAT  
GTCTTGTTGTTATAAGCTTCACCCCAGGTAGCAAAAACTATTCTACTCA  
AAAGGGGTAGACATATGTTAGTTCTCAAGATCATCTTTGGTTTCAGAGT  
TTAACTCAAGTGATTGGCATAGGCTGAATCCATCTCTTAAAGGATAATC  
AAATTTATGTTGAAGACTTGGTTGTCTTCTACTATGAAATGGGAAACAT  
TATCACTACTCCTCCCCTGTCAACCACCAAGTGTGGCCACCACCACCAACG  
TTAGTGAGTGACTGTGGTGATATGATGACCAAGTGGCCAGGTCAGCAAGT  
GGTGACGCTGTGTCTCACTGGAAGAGGTTAAAGTCTTTCTAAAACAAAA  
TACCATTGGCATCAAAGTGGCCAGAACTCCCTTCTTTGAGCTTTCCCTGT  
GTTAGAGCCCTTCTTGGGTTGGGAGTTAAACCCATAGTCTTACCTTCAT  
CTGTTTAGGGCCATCAGCTTCAAAGAACAAGTCATCCTCATTGCCACTGT  
AATAAAAAACAGGGACATGTCTCAATTATGTCTTCTAAACAGGTTTATTTT  
TCCTTCCCTGTGTACAAGACTTGACTGTTTATAAGAACTGCAAACAGCC  
TGCCTCTCAAAGCTGCCTGAAACACCTGGCAAGTTTACAGTGATATGCC  
CAGAACAGTCCAGAAGGCAGATTCTAGGCCTGGCAGGTGGGCACCCTGGG  
TGCTCCCTGTTGGATCTTGAGGCCTAACCTCTAGCCAGCAGAGTCAGCT  
AAAATCTCAGCTCTCCCTCTCCCTCCAAGCCACACTTTGCAAAGGGATTCT  
CTTGATTTGTGGGCTTGGAATCTTTTCTCCCCATTTGCCTCTGCAGGAAG

FIG. 3 (43 of 52)

45/118

CCCTTGCAACAACACA TGGATAGCCTCCAGGTCCCAAGGCTGGAG  
CTTGTAATGGGAAAGTAGTCTTTAAATCAGATTTACTTGGCACCTGT  
GCCACTGAAAGAGGCAATTTAGGGGAAAAATCTGGTCTCCAAGCACAGAT  
AACACTCTACTCTTGAAAGAGGAGACCTGCTCATGTTACTGGTCTCAGCG  
TCTCCACTGACCTGTAATAAGCCATCATTTCACTGGCGAGCTCAGGTACT  
TCTGCCATGGCTGCTTCAGACACCTGTGTAAAAAGGAGAAAATGAGTGAC  
TTCCCATGACGGCTACGTTTCATGTGTGATTCTCTCAGCATCCAGTGCA  
TGGCAGTCATGCAAAGAAATGATCTCTGAGTAAATGAATGAATGTGTGAA  
AGAGAAGTCCTTTGGGTCTAGAGAAAAGCATTGTCTAAACCAAAACCCAA  
CTAGCAATGTATTGGCTAGGAGAGCTGGAGCAGAGGCTTTGACACTAACC  
TTTAGGGTGTGAGCTGTTAGATAAGCAGTATCCATTCCCAGAATATTTCC  
CGAGTCATAAGCATTATATTACACCTGGCATTTTTGCAAAAAGCTGAGAG  
AGGGAGGCAGAGAGGGAAGGAGAGGGAGAGACAGAGAAAAGAAAGAGAGAG  
AGAGAGAGAATATGCATACACACAAAGAGGCAGAGAGACAGAGAGACTCC  
CTTAGCACCTAGTTGTAAGGAAGATTAAAGTCATACTTGAGCAATGAAGA  
TTGGCTGAAGAGAATCCCAGAGCAGCCTGTTGTGCCTTGTGCCTCGAAGA  
GGTTTGGTATCTGCCAGTTTCTCCCTCGCTGTTTTTATAGCTTCAAAAG  
CAGAAGTAGGAGGCTGAGAAATTTCTCTGTTGAATACCTGATTTACAAT  
CAAGTTAAAGGAAAGGGGAAAAGAGTATTGGTGGAAGCTTCTTAGGGGAG  
GGGACTAATAAACTGAGATAATTCTCTGGTTCATGGAAGGGCAAGGAGTA  
GCAAACTATGACACATTTTGCAAATGTATCACCATGCAAATATGCAATTGT  
TTTCCTGACAATCGTTGTGCGAGTTGATGTCCACATTAAATACTGGATT  
TCCCACGTTAGAAGAATGTTTAAATTTAGTATATGTGGGACAAAGTGGA  
GACACACAGATTTATACATGCACATACTTTTCTTCATTCACTTCTTTGTA  
CTTAAGTTTAGGAATCTTCCCACTTACAGATGGATAAATGGGTACAATGA  
AGGGCCAATAGCCCTCCCTGTCTGTATTGAGGGTGTGGGTCTCTACCTTG  
GGTGCTGTTCTCTGCCTCGGGAGCTCTCTGTCAATTGCAGGAGCCTCTGA  
GGAGAAAATTGACCTTTCTTGGCTGGGGCAGAGAACATACGGTATGCAGG  
GTTCAGGCTCCTGACGGAGTTGGGGCAACCCTGGAGATAAGCTCACACAA  
CCCTGCAAGACCAGGTGCTGTTACCTAGCCAATCTCATGGATGAACCAG  
ATCAATGCCAGATGAGCTCTGCCTAAAAATGATTTTTTGGTGAACCTCTGAA  
AAGTGGAATATTGTTTCTGTAAGAATATCCATCTGAGACTCTATCTCTTG  
GTAATACCAAGAGTTATCAGTTTCTCTTTAACCGAGACACCAGCAAAGTG  
CCTGCTCCAGGGTAATGCCCAGGGGAGCCCTCCATTTGTAGAATGAATGA  
GAGTCCAGGTTATGAACAGTGCCTGGAGTGTAGGAACACCCTCCTTTGCC  
TCTTTGACAGGTCTGCATCATAACACTTTTTTTTTTTTTTGAGACAGAG  
TCTCACTCTGTGCGCCAGGCTGGAGTGCAGTGGCAGCATCTCGGCCCT  
GCAAGTTCCGCTCCCGGGTTACACCATTCTCCTGCCTCAGCCTCCCCA  
GCAGCTGGACTACAGGCACCTGCCGCCACGCCCGCTAATTTTTTGAT  
TTTTAGTAGAGACAGGGTTTACCATGTTAGCCAGGATGGTCTCGATCTC  
CTGACCTTGTGATCTGCCCGCCTCGGCCTCCCAAAGTGTGGGATTACAG  
GCGTGAGCCACCGTGTCCAGCCTGTAACACTTCTTATAGCACTGAGTTGA  
AACCTTGCTCCTCCTGGTTCCTCCAGGAACTGAAATCTTTTTGAGCCAA  
GTCTAGCACAGTGCCTGGCATGTACATTCAGGTGGTAGAGTTTGCTGCTT  
GAATGGGTGAATGGGAATTTGACAGCATTTTTATTCAAATTAGTATGTGC  
CAGGTATCGTGCTCGCTCTGCATTATCCAAGGGAGTGAGCCTCTGTGCAA  
GTATTTGAGACACGAGGGAAATAGGTTCTACTGTGGGAAAAAGAGCATTT  
CATGGACTTGCTCTCCAAGCAGCCTTCTGATTTTTAATTGGCTCCAGT  
ATCTTGATATCAGGAGTCAGTCACAAGAACTCCATCTTTAGTAAGTTATA  
TTTTCCACAGGAAATCTAAAGCTGTTCAACATGTTAGTTTCTGTGAAT  
TTGATAAGCCATAATCCATTCTAACACTGAGCCCTCCTGAAATTTGGTG  
TCTGGTCTGCAGATAGCTAAAAGCCCTGTCTGGGTGGCCTAGGGACTCC  
TCTGTTTGCCTCCACAGGATCCACTTTGCAAATTAACCACTGGTTCTCC  
CGTTGTAGGAACTGCCACCTTCTCAGAGCCTGTCTTCTTCTTCTTCTTCT  
CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
TCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
CTCCCTCCCTCCCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
CTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
CTCCCTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT

FIG. 3 (44 of 52)

46/118

TCTACCTTTATCCCCC...GCTGGAGTGCAGTGGTACAATCATGCAATCA  
TGCATGATCACAGCAGCCTCAAACCTTCTCAGAGTCTTTATGCGGCAA  
CCAGCAGGGTCTGGAGGGTTGGTGGCTCTGTGAACCTCTCCTGACAGAACA  
CAGAGATGTCTTTGGTCTGTTGATGTGATTACAAGCTGAACGAAGGAAGA  
TCAAAGCCAGTGACAGGAAGGGAGATATGCAAGGGACCCGAGCATCAGCT  
CTGAGTTAGTCCATTCTGCTTCTGGGACTTGGGATACAGGTGAGAAACCT  
TGAGCTTCTACTTCTCCATCTTCCAATTGTAGCATCCAGGACCTCAGAAT  
CTGCCAGCTAAGAGGAGCCCTAATGATTGTCTGGTGGGATATGGTGGGAC  
CACAGAGATGAAGACATGAATAGCTATTTGAATGTGAACAGCAGACGAAG  
AAATCAAGGCTAGGAGGGTGAAGTGACTCATCCAATAGCACAGTGTGGT  
TGAAGCAGCACTAGTATCCAGGTTGCATGAGCCCCCTGATGCTTTCGCTCG  
AGGGAATTTTGGAGCCATGGGGCAATGCCCCCTGACGTAACAGTCTCCA  
CAGTTCTGCCATGTCTCATCTCGCCCTGTAACTGGACCCAAATCTGCT  
ACCATCCCATCCATCTCAGGAAGTGAAACCTCTTATGTCAAATAGGTTGT  
GCAACGTATGTATCAGATCCTGTCTTCCCAAGGAGACCGCTCAGGCCACA  
GCACTTCTTCCGATCCCCAATGAGCAGAAAATATCTCGCTATAAACATA  
GTTGGCACTAAGGGAGGGAGTGAAGAGTGATGATGATGTAGATGGTGTAT  
GTAGCCCCAAGGAAGTGAACAAGCAGAGATGGGGAGCTGGAAATGCCAG  
GATGCTCCAGCTTTTGGGGAATTATTCAGCTCTTGAGTCACTAAAGCCTT  
TCTCAGCTGCAAGTTCCTCTTTACCCTGTGAGGTCACTTCTTCCAAGACAG  
GAGACTGACATTTATTCAAAGCAGCAAGTGCCCTGATACCATCTTGTGTC  
TAATCATGGGCTTCGCAGCCAGTTATCAAGGTTGATCTCATCTCATTGGT  
CTTCAATCATTTTGAACAAGAAGACAAGCAAATAATCATGGGTTAGTTTC  
TTATATTATTGTGTGTACATGCAGTGATGTCTGTTCTTTGTAGTGAGCTG  
TTCCTTCTTGTTCACCCCTCTTGCTTAGAACAGAACTAAGCAATCTGCCC  
CCAACATTTTCCCCAATTTCCCATCTCATTCTTGGCACTGGCTTCCTAAT  
ATTTGTTCTTATGAGTCATTTTCTGTATCATTTCCATGAGTCCCTCTGG  
GATCTTAAAGTATGAAAAATGTTGTGTGTACCCACACCTGTCTTTGTGGA  
TATTTCTCTCCTTTCCCTTCTGCTTCTGGGATTATTTGGGAATGGGCACT  
ATGATTTTATCATATCGCTTCCACTTCTTTATGGCATCATCTCCAATG  
GGCTTCTTCTCCCTCTTGGATCCAGGTTCTCAGATTGGGGACATGCAGAG  
TCCAAGGAACATTCCATTCTCCTCCCTGGTCTAGAACAAGGAGGGCTTAG  
ATATATTAGCAGGTGGCTGGGGCTGGCGAGCTATGTAGTCTCCAATGGCT  
TTTCCCTGATGTGGAGTTGTTATGTCAGTTCTGGGAGACCAATAAGACC  
TTGTCTTCTTGGATCCATCAGAAAAAGCCCCTGGGTGGGTAAAGATGG  
ATGGCAGGGCTCTCCTACTCTATGTCTTTTCTCACACCTAGTGGGTATAA  
GAGAGGGGACCACAAACAGAGGGGGCTCTGGTACCACTTATCCAGGGTCT  
GGAAACATTTTCTGTAAAGGGCCAGATAATAAATGTTTCAGGTACAATA  
CTCAACCTTGCACTATTTCAAGAAAGCAGTCAGATAATACATAAATGAAT  
GGGTGTGGCTGGACTTGTCTGCGGTCCCCTGTCTTATATCATTGTATTA  
TATCATTTTCTTACATACAAATTTAGAAGCAATACTTAAAAAAAAAAAA  
GCCGCTCTTTATTGAGCACCTACTAAGTGCCAGGTACCTTTTTTTTCCCTC  
ATTATCTTATTAACCTCTTATAATAACCTTTAAAGTAGATAATATTGAAC  
CATTTGACCTATGCAGAACTGAGGTTGAGACAATAAATTATTTAAGACC  
GCACAAACAGTAAATGCTGGAACACGACTCAAATATGGGTTAACTGAAC  
CAAAACCAGATCTTTATTTCTCACTTTTAATTGTTACATATGTTTATTGC  
CTCATCTCCTGTCCACATGGTGCCCATCGGCAGACTCCTTTCTCATTCTC  
AGTGATTGAGTGACATTCTAAACTACATTGGCCTGGCAGATTACCTCTG  
TCCCCTAAATGTTTCCACATTGTCTTTTAGGATTGAGATCCTCTCTGTT  
CGCTTGTCTCCCTCCTTCTTCTTCTGGCGGTGACGTGCTGTGTGAATT  
TGTTCTTTCTCCTCTCAGGGTAGTACTGGGACTTCCAAATCAGGGTTT  
TTAGTGATCTCTCTTCCCTTTTCTGAGTTTCTTCTTATTCCCATTCAT  
TTCTCATCTATAAGTGGCAGCTTTGTTGCTGGAGGATTTCTTTGTCTT  
TTATTCTTCTTTAAGACTTTGTCTAACTGTCAAAGCAATCCCTTGAAG  
GTATCTGTCTTGAATGTTGTGCTTATGATGCTGAAAAATACTCTCTTC  
CTAAAGCTATTATAAATGCT  
>Contig42  
GGCTAGCTGCAACTCTTGAATACAAACACATTGACATGCACACACTTT  
CTGGCTCCCAAAAAGAAAAAATCAATTATAATAATTCTGATCCT  
TTGCTTATTTCCACAACTCCATGAAAATTGTACATTGTCCAAGCAACAT

TTCTTAATATTCTCTT. TCTCTCATATCCATTTTCCTTACTGCTGTC. C  
CACCTATCTCTTCCAACTCCCTGTTAAATCCCTGCCCCAGCGAACTTT  
TATTCAATTTTGTGGAATGGAGGCTGCACTGATTTAAATTAAAAAAAAAA  
AAAAAATCCCTACTCCATGTCCCAGATCCCTAGTTGTTTTTTGTTTTTGT  
TTTTCTGTAGACAGGGTCTTGTGTCTTCCATGCTGGAGTGCAGTGGCATG  
ATCATGGCTCACTGCAGCCTCAACCTGCTGGGCTCAAGTAATTCTCTTGC  
CTCAGCCTCCCCAGTAGCTGGGAGTTTCAAGTATGTGCTACCATGCCTAGC  
TAATTTTTTTCTTTTATTTTGTAGAGACACGGTCTTGCCAGGTGCCCCAG  
GCTGGTCTAGAACCCCTGGGCGGACGTGATCCGCCTGCCTCGGCCTCCCA  
AAGTGCTGGGATTACAGGCGTGAGCCACTGCTCCCGGCCTTGGGTGCAA  
TTTGAGCTTTCTCACTTATTAGTGTAAGACATACAGCTAATTTCTAAATC  
TTCCAAACCTCAGATTTTTCATCCATGAAGTGAGGATTATTATAGAGCTC  
ACTAATAACATGGCTTCAAAAATATATAATGCCAAAATTGAGATCAAAT  
AATAAATCTATATTACATGGGAGATCTTAATGTACCTCTTATATTATTGA  
TAGACTAAGATGATCAAAAAAATAGAAAGAGAGCAGTAAGGAGAGCAAGC  
ATTTAATCAATAGGACCAATAACATTTTAATCAATAGGATCCTCAGGAATA  
TATACAGAATACCAAACCTAACAACCTGCAGAAAACATGCCAAACATTTAG  
GTACAGACATTGTTGGAAAATGCAATCTTGAAACGAGTGGACTGACATTC  
AGAAGATATTAATAAGAGCACTAATGATGGGGATTGCAACCATGTCTTTA  
CTGACTTCCAGAAGCTTCTTACAGTAAACATGAAATCACATAATTTCTTC  
CACTTTCCTAGTGTCTTGTCTGGGCTCTGTCTGCTTACTGTETAAT  
ATCTTGGCCCCCTTAAAGTTGCTAATCTTCCAAACCTCATTCTGTGACT  
GGGCGCTGGTCTTGTTCATGGGCTTGAAAATACTGACTGTACACTTA  
TCTGGAGCATCCAGTGCCTACCACCTGACCCAGATTCTTCATTGCGCTCC  
TCCCTCCTCCACCTATTGGAATTGCTCATACCCGTGTGAGACCCCTCCC  
TTTCCCCCATCTGAATTTTTATCAAGACAACGCACTGCCATACTCCCTC  
GTACCCTGCTCTGGGCATCAGACTGAATGTTTGTTCATTGAGGATCTG  
CAGCTGCATCAGTTTCCCCAGCACCGTCCAACCCCTTGAGCATGGCTAGT  
CCTAAAGCAGAGAATTAGCCTTCTATCCCTGCTGCTATACATGCTGGGA  
CAAATAATAAGAAATGACAGCATTTTATGATAATGCAGGCTGCAGGAGGC  
AGGAGGCAGGAATCAAATTCGTGCTTATCAAATAGTGCTCCAATTCTTTG  
AATATTGGACTATAGAATATGTCATGGATCTATGCTCAGGTGGGTTCCCT  
ATTACTCACTCCACTGAGGCCAGTTGTGGGATTAGCTGTCCAAGAGGGA  
GTTTCAGTCTCACAGCATAGGCTCATTCTGAGAATTACTGGCCCACTT  
GTGTGGAGACCTCCAGAGAACAGAATCTGGGTTGGTGCCATGTACTTCCA  
GGAGGAGAGAAGTGGCAGGATGCCCAGCCCCACAATCAGAGGGGAAGGGG  
CAGAGCCACATGTATGAAGATCCTCTCCCCAGTACGTGCCAATCACAGGG  
CTTCCTAGCTTTTGGGCCAAGGAAACAATGTGGGAAGCAAAAAGGACAA  
TTTTCTCTCCCTTTGCATGAAGACTGAGCAGTTTACCAGATTCCCAGG  
GAAACACCCCTTCCACTCTGGGTTGAATGTGAGTGAGAGACATTCAGCTGG  
AACACTAGAAAACTATTCTCTGAGCCACTCACCTTTAGCCCTAGAAAGT  
GTTGGATTGTCTTTCATCTTTGCCACAGTAGAGACTGCTGATAGCATCA  
GAACATTGGGCTCTGGAATTAGACAGATATGGGTACAAATCTGAGCTCTCT  
CACTTATTAGTGTGGGATGTAGAGCAACTTTTAAATCCTTCCAAACCTC  
AGACTTCTCATGCATGATGTGAGGATTGTAATAGGGCCACCTAATAGGG  
GTTTTTGAGAATTAAAAAAGTTATTCAATGAACAGCATTTAGCAAGATGC  
CTGACCATTGAGAAAATAACAAATTGTTTATTATTATTGTTATTATTAAA  
CATCTTTCCTGCACCTTCTGACTGGGGGCATCGTATCATCAGAAATACTT  
AGGATGGGATGGATTCTGCTGAGGCTGAGTCAAGGGTGCAATAATGGAG  
GAGTGAAGAAGGAAGAAATGGAGGCAGAAATCCCCAGGAGCCCAGCATGG  
TACAAGGCTGAGCTAGTGCTGCAGAGCCTCCTTGGAACAGCCACAGAGCT  
TGCATCTGGCCCTGGGAGGAACCTCTTCTAGCTGGCAGGACCAGCCACAA  
CAGTGGCCAGGGGATTTCCCAGGGCGTGGGCTCCTAGGAGTTTCAATTTGGA  
CCAAGCCTGCCTGGAGAGGGGTTATAACAGGGATCCTTCCCTACTGGCAG  
GTGATTTACCCCTCGGTGAGAAGCTCAGGCATTGTTTGTGATGGAAGGTGG  
AAGGCCCTGTGCTGGGCCAGTGACTATCAGGGATGGGCGGGTGGCTGGAA  
AATAGCAAATAAGACAATATGATAACACAGTTAACCACCACACTATGTGA  
AGCTACAATATGGGTATCTGTAATAGACAATTCCAATGTAGAGAATAATT  
CTAAGGTGTCAATCTCCCCGCCAATGCCATAAGCACACGGCCTCTGCCTG  
GGTTTCTCACTGTGGAATGTCTCTGGTCTCCTCATGCCCAGAGAGTGG

GAAGTACTCCTACTTT. .CACC GGCTTT CCTGT CATCTCCCTGCAGCC...  
CCTCAGCCCCCTCTGCACAGGGAGGTTTCTCCCTGCTGCTGCAGTGCTT  
TGTACTTGTAGTGGTACCTGCACACAGGTATTGGTGTCTTGTCTCACC  
ACCTTACATCACTGTAAGCTCCCCAGGAGCAGGCTTCTGTTTGACTCAC  
CTGTGATCCTCCACCTCCCACCCTGTAGTGCCTCAAGCATTGAGGACAAT  
CACTGGCTGCCCCCTTAACCCAGAAATGCTGCCGAGACAGGAGGCCATGGC  
CCAAGTTCCTGGAATGGGGTATTACTATGTCAGCACAAAGGCCTTTGCAC  
AAATGAAGGCTTTAAAAATGCAGTCTTAGTCAGGTGGAGGAGGGCTTATA  
GGATTCCCAGGAATCTGGATCATTCTCTTGAGAGCTTTCCCTTGTCTCTG  
TTAAACTCACATCCTACGGCCCCAAATAACAACAAAAATGGATGTAAAT  
TCTTGAAATAACTTGTGGATGGGGGAACAAGGCCACCCCCCAGATCTGC  
CAGAAGCTTCAGGTGAGGGTCCCAAATGCCAAAAAGTCTGGTATCAGAGA  
GGATGGCCAGTGACCTGGGGACACATGCCCTTGTGTGTCACTCAAGGA  
GCAGCAGCCTCGGCCCCGCACAGTGACCAGGACCCTGGCTTCCCACGCTG  
GGCAGGAGCTGGTGTCTGATGAAGGGAATGCCTGGCAGCACGTGCTGTCT  
GTCTCCTCGTGTCACTTACCTGGCTTTGCTGCGAAGAGGCCACTCGCAT  
TTCTCAATTTTTTATATTTTTTTAATTTTTTTAAATTTTTTATTTTT  
TATTTTTATTTATTTATTTATTTTAAATTTTTTTTTAAATTTTTTAAATTA  
TGCTTTAAGTTTTAGGGTACATGTGCACATTGTGCAGGTAGTTACATAC  
GCATACATGCGCCATGCTGGTGCCTGCACCCACTAACTCGTCATCTAGC  
ATTAGGTATATCTCCAGTGGTATCCCTCCCCCTCCCCCACCACAA  
CAGTCCCCAGAAATGTGATGTTCCCTTCTGTGTCCATGTGATCTCATTG  
AATTTCTTTAAAGGTGGAATCTCTCAGTGGGGTCTAATCTGTTCAAGAAAT  
ATCAAAAGAGTATCCTTGGGAATGACTGGAATTCAGAGTCATCTGGTAA  
TCCTCATAAAACAACCTCCTGGATGTCTCTCAGCACATCTCCACCTTGAA  
CGCAGGAGGCTGGTTCAAATGGAGGAGCATCGCTCTACTGCACCTTTTTTT  
TTTTTTTGGCCTAAAGTGCAAAAGGGGATACGTTTCATGTAAATAAATCA  
ACTGCAAAATCGCTAGTTATGCTGAGCCCTGTCCCGTGTGTGGACACAAA  
GGAACCAAAGGCTTTTCTCCCCGCCAACACACACATAACACACACACAA  
AATCATAAAAAACATACATACCCCCAACACATAACAACACACACACAC  
ACAAAATATATACACACAACACACACCAAAACATGCCCAAAACCTGTGTC  
CAGAGATAGATCCTACTGGTGGGTTTGTGGTCTCGCTGACTTCAAGAATG  
AAGCCGTGGACCTTCGCAGTGAGTGTACAGCTCTTAAAGATGGCATGGA  
TCCAAAGAGTGAGCAGTAGCAACGTTTACTGTGAAGAGCAAAAGGACAAA  
GCTTCCACAACCCAGAAGGGGACCCAGCAGGGTTGCTGGTTGGGGTGGC  
CAGCTTTTACTTCCTTTTGGCCCCCTCCCATGTTCTGTTTCCATCCTATCA  
GAGTGCCCTTTTTTCAATCCTCCCTGTGATTGGCTACTTTTGAATCCTG  
CTGATTGGTGCATTTTACAGAGTGCTGATTGGTGGCTTTTACAATCCCCCT  
TGTAAGACAGAAAAGTTCCTGATTGGTGTGTTTTACAATCCTCTTGTAAG  
ACAGAAAAGTTCCCCAAGTCCCCACTGGACCCAGGAAGTCCACGTGGCCT  
CACCTTTCAACTCCATAATGGCATGAAAATACATATGTTGTACAAAACAT  
ACATACACAAAGTATACATGCATCTCCCCAAATATACACATACCACAGAA  
ACATACACACAGGAACTCAGCTACCTGTCAAAAGTCTGCATGGTGATTGC  
CTCTGCAGTGAGTAGTTAGAAAAGTGAATTTGTTTTTCAATAAATTGGAG  
TCCTTAAAAATCGTTGTAAGATAGAAAATTTTTTAAAGTATATAAAATAA  
AATATGTATGTCCTTTGGTCTAGCATTTACACATGTAGGAATTTATCCTA  
GTGGAGTAATCAATGATATATGCAAAGATTTGGACAAGCATATTAAGCAC  
AGAATTATGTATGCATATGTGTGTATATATATATATATCTCATAACATA  
TAATAATGTAAAAGTGAAAATAACTCAGATGTTCAAAATTGAGGATTAGT  
TAGACTATGATCTGTCCATATGTGACATAACAAGTTAGCTGCCCTTATTC  
TCTCGAGCTTCAACCTCCTATAAACAGTGTCCCTGTATATCAGTATTGG  
TACAGATAATCGAACTTATTGAGGTTTTTACATGGGGCAATAAAGGCAAGA  
GTTTATGAATACTCCATACTACACTAGGTAGCACCCCTATTAAAGACAA  
ACTCTTCTCTCTCATTTCCCTTCCCTTCCGGAACCACTTGGTTGAATCTC  
TACAAGTCTCTATTGCAACTGCCTCAACATGGCACCCCTCCCTGCATCTCC  
ATCTTCCCTGTCTGAGAGCAATGGCCTGCTGCCCCCAGCTCACATCCT  
CATTCATTCCAGAAGTGAGCACCAAGTGCCTACAGTTACCCCAACC  
ACCTTCTTAGAAGATAAGTTAGTGTGTTTGTGTTTTGACTTTTTTAAATTTT  
CTTCTCTTTTCTTCAATCTCATCCCATCCCAAGAGGTTTATCAAGA  
AGTTCTCTAAAGATATGTGTCTCTTATGGAATTTAACAGAAATCAGGGA

ATGTAATCTAGCCATC AGGGAATAACATTTTTCCAGGTCTTTAGAC  
ATAATGGAATACCTTGCAGTAATTAGATACACTATTGTAGAAAAGTATTG  
ATGAAATGGAACGATGTTTGAGATATCATATTGAGTAGAAAAGGCAAGAT  
ACATTAAGTAGGAAATGTATCTTACAAAATAATTTGTGAGACACACTCCT  
ATATTTGTATGTTATATAAATGCGTATGTGAAGAAAGGCTAGAGGATGAG  
ACCACAGTCTTCGGTGAAGTTTAAGAGATGATGCTGCAGCATGCTCAGAA  
AGGCTTGGTATAGTTTTTTCCAGTAATTAAGGACTGATCTTAGGTAAATT  
GTCCATCCTCTCTAAACTGCACCACCTTTTGTCTGTAAAACAGGAAGGAT  
GGTATTTACCCCCAGGGTCATCAAAGGATTTGGTTGGAGAAAAATAAATA  
AATGGGCTGAGCCAGACCTGGCACAGTGAGAGCACAGTGGTTGACTATT  
GTGCTGGCCTGTTGTTCTGTGTTATTGACATGCTGCTGGTGGTGGTCCA  
GAAGCTATTACCTTAATTGGTTATGTGGATTTCCCCTCATACTGAGCAGC  
TGTGTGTTGGTGTGTGTAAGCATAGCCATACACAGTAAGTACAAGGGCAA  
ATGTGATGGAATAATGCAAGGAAGTGCAGATAAATAGCTAATGGGCTGTA  
GAAGGAAGCTAGTCTTGGAGGGCTTGATCAAGGAAGGTCTTTTGCATG  
TCACCTTTGAAGAAGAGGGGACATAGAAGAGGTATAGTGCATCCCGGAGT  
GTACCTGGAAGGGAACATGAAAAGAGGACATTTTTCTCTGGGACATGGGG  
ACTCCACTTGCTGAACTCTGGAATTGGGGCAAAGAACCATCATGAGAAC  
AAGGGCTTCCTTGAACCTCCCAGGCTCATTGGCTGATCTAAACCCTGTGT  
CCCCTCTTTCTTCACTCTCCTCTGTTTTCTATACCTGTATTATTGGACT  
GGACTGGAAGCCACCTGATCTATCACAAGTACCTTGAAATGTGTTGAATA  
GGTGTGGCAGTCTTGTAGCAGAGTGGCACTACCCACAGGAATTTGTT  
TATACCTTTGGCATGGAAAATAGCAGGAAATGAGTGATCACTGATAACTG  
AGGATGCTATTTATTATTGGCCAAAGGAATACTTGTGTTGTATTTGCATA  
ACCACTCACAACTGTTGATTACAAATGAGTACCAGACCTAGCTCCTTCA  
AGTAAAGGATCTTGAGAACTGAAGGCAAACAGAGCTCCAGGAGTCCAAGA  
CAGAGCCACAGACCACGAGGATCCCTGGCCAGGTAGGTGGTCTCCTGC  
ACTGGCTTTCAAGGCCAACAGGATGGATGGGGAAGTAGAGTAGCATCTGG  
CCATCTAGACCCTTGCTTTTTATCCCCACTGGAAGCACATCTGAATTTCT  
AAATATGATCTCTGAGACCTGCCAGAACACCTTGCTCTCAGCCCCAGTA  
GCAGCCTGCTCTCTCCAGGAGGGCTTCCACTAACAAGTAGGGCATTTGCT  
GGAGGGCCAGGCAGACACTAGCTTAGGAAATCCACCAACCCTGGAAATGC  
TAGTCCCTTCTCTGAAGGCTCAGAAGACTGACTTTAGAGTCTAGAAAATA  
TTGGTCTTGGAACAGATTTTGAGTGCAAAGAGATGGACTTCAGATGGC  
CAGATGCACTGCTTCTTTAGGGAATTCTGTGAAAGCTCCCTGCATTTATC  
TTAATACAGGCAGCAGATTTTCATGAGTACCCCGAGGGATGGCCCCAGGT  
CCTCCAGCCTGTGAGCATCCTTCTGTCTTTCAGCAGCACCACAGTATCTT  
TATATGTCTTTGGATACCTACGTTTCTGCCAGACATCTCTTGCTCTGATG  
TTCTGGCTSCCAAATTCTCTGTCAAGCGCCTCCAATTTTTGTGTCTTT  
GATTTACCCCAACATGACAAAGGCAGTTGTGCTTCATGTATTCAGGGATA  
CTGCCAAACCACAAACAGGTTAAATCAAATAGCAGATATCCCTGTTCT  
AAAGACCCATCAGCTCTACCCACCTGCTCCTGCTCACCGTCCTTATTGTT  
GAGTCTGAAGCCCTTCTTGTCAATTTTTATTTTTTGCATGAACAATTTA  
GTTCCCTTTGTCTCACTCCTAAACCTTTCTCAAAGGATTGGATTTGTACA  
CAAACCTGCCTATCTCTGCAATCTTAGAAGTGATATGATTCTGAACAAATC  
ACTTAACTTTTGATTTTTTATTGGTAAGATGGGAATACCAATTTTTGCTC  
CACTTCTGTCTATGTTGGCCTGGGCTGATGTTGAAAGCTCTCGGTCAAC  
TGAGATAGGGTGTGCAGAATTTATATATATAAATATATCTCCTCCAACCC  
CTCCCAATGAAGCAAGTCAGTGAGTCAATCCTACCCTAAGATATTAGGG  
ATTGAGCCTCCTGGGACATTTGGTGGCTTAGGTTTTCATGAAAAGAGGTT  
GCAGAGCAACTGCTTTTTGTTAGGCAAAGATTAGGCTACTGCAGAGACTC  
AGCAAACCTTCTATAGAAGGTGTCAGATGGTAAGTATTTTAGGCTTTGCTT  
GCCAGATGATCTCTCAACTAGTTAACCATGCTATTGTAGCCTCGAAGCAG  
CCAGAGACGATCTGTAAACAAGAGCATGTAGTGTGGCATAAATATAGTA  
CCGCG  
>Contig43  
GCAATAAGTCTATTTACTGTAAAGTTAATCAAATTTACATTTTCAGAACAC  
TTAATCTGCAAGAGTCCTTTCCAAGACCCTATACCTAATTTTGTGTTTAC  
AATTTTATATTTGTTTTCTTAAAGAAGACCACCAATATAAACTATATCCA  
GCCTTCATGATAAGTACATAAGAACTATGCAATAAGGGGGAAAAAAA

CAAAGAAAAATACCTAC . CTACTAATGGTTCACCTTCTGAATAGCAT...  
TCATAATGATACAAGCACTCATTACTAGTCTAGGAAAATGAAGATATAAT  
TGCATTAGGAAGATCAAGAGGTAGGAAAATGTGGATGTGTGTGGTATAGAC  
TAGGGCAGGACAAAGAACCTAAATCCTCATTTTCTAAAGATAATTGTAA  
TACGTAAACTCAAAATTCAAGAAGTAACAGTAAAAGCGGTCATTAAGAA  
ACAAGCACTAAACACCAGATAGGAAGCGAGAGATGGGGGAAGAGGGCAAG  
AATCTGATTATTTTTTGTCAACAAATTTGTAAAACCATTTGACTGTTTAC  
ATGTAGAACTTGGATCTTTTTTAAAAAACACAAAATAATAACTATTAT  
TTTTTAACTGGATTTTTTGAAGAAAGATAAAAGTCTCATTTTAGTAATT  
AAAACCTCATTCCAGGTTAGTCCACTCAAACTTATATTGAAAATTAAAA  
CTTTGGGAGGCTGAGGCAGGCAGATCACCTGAGGTTGGGAGTTCGAGACC  
AGCCTGACCAACACGGAGAAACCCGCTCTCTACTAAAAATACAAAATTAG  
CTGGGCGTTGTGCATGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAG  
GAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACA  
CCATTGCACTCCAGCCTGGGCAACAAGAGTGAACTCCATCTCAAAAAAA  
AAAAAAGAAAAAATTAAAAACCTCTGGAAGTTGAGTTTGCAGATATTCAT  
TATGCTCATTTTTAACTTGTATGTTTGGAAAATGTCATGATGAGAATTGA  
GGTTGGGGGATGAGAAAAAAGAAAAACATCAACCCACAGCCCATTCAA  
TTTTAGCCCGACCCACAGCTCCGGGGAAGGGCAGCAGGTCCATCCTTCA  
CTCTTTCTTCACCTCTTTCCCTCCTTCTGGCTCTTCCACCTCTAAGTTG  
GAGCCCAAGAGAGGCACTGGGAAATGGAAAAGTCTTTTGTACGTGGTAC  
TTGCCGGGGAAGCTGCCATGAAGACCTGGCCCCACGGTGGGGAGGGAATG  
CCCAGCTGAGGCCTCGTGCCCATGCTAGGATAGACTCGTCCAGACATGTC  
AGGTGGTCTGACAGGGCAAGCAGCAGGAAGTCATGTATGAGTATGAACTG  
ATCTGTATGCAAGGGCGGGGAGAACACGCGGAGGAATGGGGCGTGAGAAA  
ACAGCACAGTACGTTTTCTTTAGCAGCTGTCTCTGCTCAGCCATGGGAGTC  
ACCAGAGAAAGAGGCTTGGAGGCGTTATTTTCACTGTGAGATGTGAGTGT  
AAAAAAGTGCCCAAGACACAGTGAGTACCAGGGAGATGCCCTCTTTCCCT  
ACCCGAATGCAGAATGGCCACAGGCCTTAAAACACACACATGGTTCCCTCA  
GAGGAGAGAGGCCTCCACAGTGGACACCCGCATTCTCCCTGGTCAGCAG  
CAGCAGGGCGAGTGCTGGGCCATCATGAAGCTTCAAGGCAATGAGCTCT  
CAGCAATAACAGGAACAGTGCCCTGGGGGACTGTAGCTGCAAGACCGATT  
TCATGTAAAGATGGCCTCTGAGGACTCCGAGATACACCAGGCTGAGACTAG  
CTGGCAGCTCCAAGTTCTTGGTCAGAAGAGAACAGGAAGTGGGAAATTG  
GAATTACTGTTACTACAATTCTTTTACATCCGCACAACCATGAGGTCCAG  
AGAGTCTCTCTTATTTTTTTTTTAAAGACAGGGTCTCACTCTGTGCCCCA  
GCCTAGAGTGCACCTGGTGTGATCATGGTTCAGTACAGTCTTACCTCCCA  
GGCTCAAGTGACCTCCTGCCTCAGCCTCTCAAGTGGCTGGGACAGCAGT  
TGCATGCTACCAGGCCTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  
TCGGTAGAGACTGGGTCTCTCTGTATTGCCAGGCTAGTCTCGAACTCCT  
GGGCTCAAGTGATCCTCTGGCCTCAGCCTCCCAAAGTGTGGAATTACAG  
GCATGAGACACTGCACCCAGCCAGTATAGTCTTTTAAACAGCTTTATTGAG  
GTACGGCTAACATTGAAAAAACTACACAAATGTAAAGTATGCAATTTGAT  
AATTTTGACAAATGTACACACCAGTGAAACTATCACTACAGTCAAAATAA  
TGAACATATCCATCACTCCCAATTTCTCAGCCCCCTTGGTAACCCCTCT  
CTCCCAACTCCCTGCCCCCTAACATCAGACAACTACTGATGCATTCTGTC  
TCCATAGGCTCATTTACATTTTCTAGAATTTTACATAAATAAAATGACAG  
AGTATATACTCCTTCATGTATGGCTTCTTTCAGCCCAATTATGTCAAGAT  
TCATGCTTATGGCTGTGCGTATCCTTAGCCCATCTCTTTGTCTTGCTGAG  
TAGGATACCATTGCATAGACAGACCACAGCTTGCTCATCCATTCACTCTT  
GACAACGTTGAATTGTCTCTGTTTTTTGCAATGACAAATAAGGTTGCTAT  
GTACATTCTGTATAGACATTTGTAAAGCACAGCATTTTCATTCTCTTG  
GGTAAAGACCTAAAAGTGGAAAGGCTGAGTCATATGGTAAATATATATGT  
CTAATTTTTTAAAGAACTGTCAAAGTGTACCCAAAGGGATTGTACAATT  
TTACATCCCCACCAGCAGTGATGAAAATTCCCGTACTTCCACATCCTCA  
CCAATATATGGTGTGGTCAATCTTTTTAATTTGGACATGNTAATGAGTG  
CAAAATGAGGCCAGAGTGCTGAAGTTACATTTGTATCCTTTTGGCAT  
CCAAACAGGTGTCAAGCATAGAAAAACACTTGTTCCTTGAATGGTCAG  
TCATTTACAAGTGGAATTCATTACAAACCGGTAGTTCTACTGGGTAAAC  
TATGCCTTACTGTCAACAGGCACATACACATACAGACAGACAGGAAGGCA

CAGAGACAAGGCAGAGC...TGATAAGAAGGTGACCTGGGCTCTAGCTCT  
GCCTATCACCTAGTAAAAATATTAGTTAAGTAGCCATGAGTAACTCACTTA  
ACTTACCACAGGCTCCATTTTCTTATCTGTAAAATAGGAACATTGAAACA  
GCTAATCCCCAAGGTTTGTGGATAATCAGAATTACAAAGATCAATGACAT  
TTCTATGAGAGAAACATATTTCCAAGTATTTGATGGAGTACATCAGACAC  
AAAGGAAAGGAACTGAATATTTTTGAGGTTTTTTTTTTTTTACCAAGAAA  
TTCACATTTTGTAAATTTTCAGAACTACCTCCTGAGGAAAGTGTAGCTG  
CAACCATTTAGAAATGATAGAAAACATCAATCTGTCTGATTCCAAAGCCAA  
GTTCTTGCTACAACGAGAAATGAAACAACCTGGATCCCTACAGATGCAGAG  
ACCTGGGCCCCACAAATGTGAATTCTGTTCCCCTACCGAATAGAGTTACA  
GTTCCATAATACAGTACTCCCTCACTTTTCCACAGTCTCACATTCCACAG  
TTTCAGTTACCCACAGTCAACTGCAATCCAAAAATATTAATGAAAAATTC  
CAAAAATAAACAAATTCAGAAAGTTTAAATTTGTGCTCATTCTGAGTAGCG  
TGATAAAATCTTGTGCCACCATCCACCTGTCCAGCTTATCGTTAGTCAT  
TGACATCGTCTGCTCCTGACATCCAACCATTGACATCATCATGACTCTAT  
GATCCAGGATCACCGAAGCAGATGACCTCCTTCTGACATATCATCAGGC  
CAATATCAGCCTAAACACTGCATCACTATGCCACATCAGTCACTCACT  
TCATCTCATCAAGGAGGCAATGGATCACCTCACATCATCACAAGAAGAAG  
AGTGGGTATAGAACAATAAGATAATTTTGGGGCAGGCATGGTGGCTCAGC  
CTTGTAATCCCAATACTTTGGGAGGCCAAGGCAGGAGGATCCCTTGGGCC  
CAGGCATTCAAAACCAGCCTGGGAAACATAGTGAGACCTCCTCTCTCTGC  
AAAAAAAATAAACAAATTTATCCAGATACAGTGGTGCATGCCTGTGGTC  
CCAGCTACTCAGGAGGCTAAAGTGGGAGGATCACTTGGTCCCAGGAGGTC  
GAGGCAGCAGTAAGCTGTGATCGTGCCACTGCCTCCAGCCTGGGCAATA  
AAGTGAGACCCTGTCTCAAAAAAAAAGGTAATTTTGAGAAAGAGACCAC  
ATTCATACAACCTTTTATTATAGTATATTGTTAGAATTGTTCTATTTCATT  
ACTTATTGTTGTTAATTTCTTTCTTGCCTAATTTTTTTTTTTTTTTTG  
AGTCGGAGTTTCACTCTTGTGCCCAGGCTGTAGTGCAATGAGACGATCT  
CAGCTCACCGCAAATCCCGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCA  
GCCTCCCGAGTAGCTGGGATTACAGGCGCCTGCCACCATGCCAGCTAAT  
TTTGTATTTTAGTAGAGCGGGGTTTCTCCATGTTGGTCAGGCTGGTCT  
CGAACTCCTGACCTCAGGTGAGGCTCAGCCTCCTAAAGTGCTGGGATTA  
CAGGCTTGAGCCACTGCGCCTGGCCTCTTGCCTAATTTATAAATTAAAC  
ATTGTACAGGCATGTATTAATTTATAGGAAATCATAGACATATAGAGT  
TGGGTACTATCCACAGTTTCAAGCATTCACTGAGGGGCTTGAACACGCC  
CTCCTCAGATGAGGGGGGACTACTGTCTCCTCAATCATCTTGATT  
AATCCTCAACACAAATGGTTTGGCCAGGCTTGCCTCTGGAGACAAATT  
GCTAAGGATTTAGAGGGGAAAAATGTAGTTCAGTGGGAAAGTCACCTCT  
GCTCCACTGGACAGCAACTTAAACCCAGGCCATGACAAGTAGAAAGGCC  
ACCCCACTCTCCTTCAACCTGGAGTATTCAGGAGTCAATCATATTTCA  
GGACACCAGGACCAACTGGGAAAACTGAGCTGCCTTGAGGAAAGCAA  
TCAGCTCCACAAGGGGCTTAAGAAACAAGCTCTGGGAGGAGTGGTTGGAG  
AAGAGTTGGGGACACATCAGAAATGCCATCAAATTTCTAAGGGCTACCTC  
GTGGTGTGACACCTGTGCATCTTCAAGGACATAAACAGATGGGATAAGCA  
GATGAGATTCACAGAGGACATCAAAATATTGGCTCCCCAGAAGGGAGAAC  
ATTCTAGTAACAGAGCTGCCAGCTGCAGAGTGGACTGTTTCAAAAGCA  
ACAGGTGCCCTGCCTCTTGAATCACCATCTTCAAGGAATGCAGTAGAAG  
GGACTTAACTCCTGCCCTGAAGAAAAGGTTAGGCTAGGGAAACAGCTCCA  
AAATTTTTTAAAGGAAGCAACATAGGCATCTACTGGGAGTTTTCTAAAG  
CCTTTGTTTAAATGAAACTAAAGAGCTGGGACAGGAAATGCCAAATTAAAT  
TAATAGAGCCTTGCTTTAAGACAATGCAAGTGGATGGTAATGAAGGAATG  
AGTCTTAGGCCCTTGATCAACCGTATTAAGCAATGCTGAGCATGGAGCCA  
ATTCTGTTCACTAGATTTGCTCAGAAAGGGCCAGACGAGAAGGATTTTTT  
TAAAGGCACCTACTACCAAAAAGCTGCCAAGGCGTCCAATGGAGCCCAGA  
GAGAATATGCTAACATAAAAAGTTGAACACCTCAATAAAAAGGGTAA  
AAGTAATTAATAGAAAATTACTGAAAGCTTTTTTGAAACCAAAGTAGTC  
AGCATTGGTAAAAGTCTACAAAAGTGGACACTTTCATATAATGTTGGCAG  
GAGGGTAAAAGACATAACCTTTTTGGAGGACAATTTGGCAACAGAGTAC  
CAAAAACCTTACAATTGAAGAGAACCTTTGGCCTGAGTGCAGTGGCTCACA  
CCTGTAATGCCAACACTTTGGAAGGCCAAGGTGGGAGGATTGCTTGAGCC

CAAAAGTTTGAGACCAGCCTGGGGTAACACAGTAAGACCTCGTCTCTATG  
AAAAATAAGAAAAGTTAGCTGGGCATGGTGGCATGTGCCTGTGGTCCCAA  
CTACTTGAGAGACTGAGGCAGGAGGATCGCTTGAGCCTCGGAGGTCAAGG  
CTGCTGTGAGCCATGTTTCATGCGACTGTTCTCCAGTCTGGGTGACAGAAT  
GAGACCCTGTCTACCCAGAAAAACAAGGCAAGAGAGAGAGAGAGAGAGAA  
GGAGAGAAAAGAAAAGAAAAGAAAAGAAAAGATGGAAGGAAGGAAA  
GAGAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAA  
AAGAAAAGAAAAGAGAAAAGAAAAGAAAAGAAAAGGAGAGAAAAGAAAGGA  
AGGAAGGAAAAGAAAAGCAAGCAAGCAGGAAAAGGAAGGAAGGAAGGAA  
GGAAGGAAGGAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAA  
GAAAGAAAAGAAAAGAGAAAAGAAAAGAAAAGGAGAGGAAAAGGAAA  
AGAAAAGGACAAAAGAAAAGACCTTTGAACCCTGAATTTCACTTTTAGAGA  
TTCATCTTAAGGAAATTCATTCCAATAGAAAATTTATCCCCAGGATTATCT  
AAATATTTGCTTTTATTTTCTTCTAGTAATTTTATGGTTTAACTTTCTCA  
TGTTTAAAGCCTTTAATTTATTTTGAATTTATTTTGGTATGAGAAAGTGTG  
ACCTTTTATTTTGTCTTTTACTTTTAAAAAAATGTATTACGATTATTATTTAG  
AGACAGGGTCTTGCTCTGTCAACCAGGCTAGAGTGCAGTGGTGTGATCAT  
AGCTCACTGCAGCCTTGAACCTCCTGGCCTCAAGCAATTCTCCCTCTTCAA  
CTTAGGAGTAGCTGGGACCACAGGCATGTACCACCATGCCCACTAATTT  
TTTTTATTTTTTGTAGAGACAGAGTCTTGCTTGTGGCCAGTCTTGCAAT  
GTTGTCTCAAACTCCTGGGCTCAAGTGATCCTGTGCCCCAGCCTCCCAA  
AGCACTGGGATTACACGTGTGAGCCACTGCGCCAGCTGCCTTTTTATTT  
TTAATTTTTTCAGATGCTTTGTTGGTTCCAAAATAGCACTTATTAACCCA  
CGCTTTCCCCCTCTGGTTTTAAATACTGCAAGTTTGGCTTTGAAATACAA  
CCCACCTGCCTTATTCAGGCTACATTCAAGGAAATCTGAGACCAAGAGTCT  
GAAGGCCAGTTTCTTCTTCAAAACCAGGAGGTGGTAAATGTGTCACTT  
CCACACTTTCTATCTATTTCTAAGAACTCCTTCTTTCCAACTCTGACAT  
GCCCCCTGGCTCAGGTCTATAGAAATCCCAGGGTCCACAGACAAAGCAGA  
ACTCACTTATGGGGAAATCTGGGAAATACTTATCTGTTAAACCTGCCCCA  
TATGGTGACTCAGATTGTCTAAAGCCCAAAGCATCATTTTCCACCCCAA  
CCATTTCTCTCCTCAGACTTCTCTATTTCTGTGGTCCAGAGTCAAGATCT  
TGATATTACCCTAGAGTCCCCCTTCTGCTCTCCTGCATACCCAGATGCCC  
CTCCCTCCCCAGATCCATTCTCCACCCTCCCTCCCATCAGTTTGGTGGG  
CCCACCTACCCGCTTCCCCTGGCCCAGGCTCTCCTTTTGTGCGCTTGAGCA  
GCAGACTGATCTCCAGCCTTCACTCACTTCATGTGGTAATCTGTTGTGT  
TCATCACTGTGAGAATCTTCTGCATCCCCCTCACTACTCTGCTGAAAACAC  
TCTAGTGGTTCTCTATTGCTCATTAAATGAAAGTCTAGATATTAAACGTAG  
AAGGCCCAGCACAAATTTGCCCCATGCCACCTACCTCTCTAATCTTTTCT  
CCTTACTCTGACAGACTCTCCGTCTGTCAATTTATGTATTCTTTTATTGCT  
CTCTTCTACTTTTAGTATGAACTGGATTTATGGATTTTTTTAAACATTGCT  
TTCAAGTATGGAATAAAGAATTTTATTTATTTATTTATTTATTTATTTGA  
GACTGGGTCTCACTCTGTTGCCAGGCCAGAATGCAATGGTGCAGTCATA  
TCTCACTGTAACCTCGAATTCCTAGGCTCAAGCCATCCTCCTGCCTCAGC  
CTCCTAAGTAGCTATGACTACGGGTGTGCATCACCACATCTGGCTAATGG  
AATAAAATATTACAATGCCTAATCTTAATTTTCAAAATTTTAAATTACAT  
TGTACCTAATGCCCATGCATTTACTTTTTTCACTGGGTCAATAGCCCTCA  
CTTTGGCAAAGGTCCCAGGCCCAAGGTAAGGCCTTACTTTTTCCAACTC  
ATCTTTTGAAGACATAAGTGCCTGTAAGTTGTACCACATTAGGTTCTAG  
GAATTTTTTCATCAAAGACTTTATCAGACTATTTTCTCTAAGTTGAGAAA  
GAGCTGGGGGCAGAAATATGGCACTGAATGACTGAAGAGAAGGCACTGAAA  
TCAGGCCAGAGGTTGCTGGAAAGAGCAATGAGGAACACCAGCAGCAATGA  
GGAGCCGGTGATGATTTTGGCTTCACAGGGAGGTGTGTACCACACCGATT  
TTATCTCTACGTGGATGAACCACAGCTGTGCGCTCCCTTGTCTCCAGGAC  
ATCACACTCTCCACATTTCCCTCCCATCTTCCGGCTTCTGCTTCCCGGGC  
CCTCATCTGCCCCATCCTGGGTGAACACTGGTCCGTCAACTGCTGGGCGT  
ACCTTCCCGCTCTGCACACCCTCCCTGGCCACCCCACTCTCACGGC  
TCGCACTGCAGAGGAGCCGCATCTCTAGCTCCAGCCCATCTGCCTCTTCT  
GAGCTCTAATTTATGTAGGCGACTCCTGCCGGTGTGCTCACAGGCCC  
ATCATACTTCAAAGCATTTTCCCTCAGAACACCATGTCCTGGCTGCTCC  
CTCCAGAAGATACATCTCTCAAGCACATCCCCGCGCTCTCACCTGGATG

FIG. 3 (51 of 52)

ACTGCATTACCTTCTC ACATTTGCCCTCCTTTGGATGTATATAGA.  
GTTTTAAATACAAATCTGATGTGCTTGCTCTCCTGCTTGAAACACCTCA  
AAACTGCCTTCAGGATAAACCCTGCCCCTTGACATGTTACAGGTTGCC  
ATGGCCTGGCCCTGCCCATCTCTTCAGCCTCATCTCATGCCCTTGCCCC  
TCGCTCTCTGGGCTTCTGCCTCCCTAGCCCTCCTTTAGGTTCTCTAACAC  
ACCATAGTCCTTCTAGTGTGGGGCCTCTGCAAGTGCTGTTCCCATTGCC  
TGAGACATGAATCCCTCTCCCTATCTCTACCTGCACCTTCATCTGATTAA  
TCCCTACCCCTTCTACTCATGATGTTGCTTTCTCAGGGACTCTCTCTGAC  
TTTTTAACTAATCAGGGTCTCCCCAGTATATATCTTCATAGCACTCTGT  
ATTACTCCTTTCTTAATGACCACCTGCTGTAGACTGAATGTTTGTCTTCC  
TCCAAATTCATATGTTAAACCTAGCCCCAAATGTGATAATATTTGGAG  
GAAGGCTCTTTGGGAGGCAGAGCCCTCATGAATGGGATTAGTAGCCTTAT  
AAAAGAGACCCCTGAGGGCTCCCTTGTCCTCCCTCCACCGTGTAAGGATGCA  
ACAAGAAAGTATGGTCTATGATCCAAAAGCAGACCCCTTGCCAGGTACCC  
AATATGCTGGCACTTGAACCTCCAGCCTCCAGAACTGTGAGAAATAAAT  
TTCTATTTTTTCATAAGCCACCGAGTCTATGGTATTTTGTATAGGAGCAC  
AAACAGACTGATGTGCCACCCAACCATGATTATACGTGTAATTTATGGTT  
TCTCTGCTAGTAGGGATGCACCATGGGGTTAGGAACCACGCTTTTCTTAT  
TTCCACACAGTCCCTTAGCTCTAAGCATGTTCTCTGAATCAAAGATCCCCA  
TCTTTTATGAATGAAGGAGTCAGTGAATGAATTAATGAAAGAACTGATAA  
CCCTCAATAATTATCCAGCCTTTTATACCTACTATTAACAAGCTTGCAT  
TCTACTCCAAATTTTATGGGCTTTAACTCTATTTTTTGGCCAGCCACATTT  
GACATTCCCTGAAGTAAATCTATGCTTTCCATCCTAAGTCAAGGAAGGAC  
CTGGACTAGTAGGGCCAAGAAAGGTCTAAATTCCATGGGTGGGAGAGAGA  
GACTAAATCTGAAAGGAAGAATAGATTGAGCAAAGGTGTAGAGATTGGGG  
AAGGCTGGACATTTGGAGAGAAGGAAAAGGAACTGACACTAAACCAAAC  
AGTCTCACAAACACAATCTCATCCTTCCAAAACCTCTGTGAAGTAAGAATT  
ACTATCCCAGGGCCAGGCACAGTGGCCCATGCCTGTAATCCCAGCACTTT  
GGGAGGCCAAGGTGGGTGGATCACCTGAAGTCAGGAGTTCAAGACCAACC  
TGATCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTAGCTGG  
GCATGGTGGTGCACACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGG  
AGAATCATTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCAGAGATCGTGC  
CACTGCACTCCAGCCTGGGTGACAGGGAGACTCCGTCTCAAAAAAAAAA  
AACAAAAAAAAAACAAAAAAAAAACAAAAACAAGAATTACTATCCCAG  
TTTTGCAGATGAGGCAATGGAAGCTCTAAAAAGTTAAGTAGGAGAAACAA  
ACATGAAATGTATGTCTTATGCTTTTCTCATCCTATTTCTCAGCCTGG  
AATGTCCATTCTCCCTCCACTATGCAAATCTAACTCTTCAAGCTAACACA  
TAGCAATGTCTGAGAAACCGTCCCTGTGTTCACTCTGTTAGCCTCACTTG  
CTCCCTCCCCATCCCTCTGTTTCTTTCTGTTATAACACTTCTCTATTCT  
GCTGGCATCACAGTCATCTCCACCTGCCTTCTCACAAGTTAAAAGCTTG  
TTAAGGGCAAGTGGTGTCTTTGCCACCTCATTCCCCAGGGCTTCTAACA  
CAGTGCCTCATGCATGACAGAGTTGTAAAACAGGTTACCAAGCTGGCTTC  
AGGCAGGTTTGCATGGAACCTGTGCTTTACAGGAATACCTGCTCCCCCAG  
GCCCTGGGTCTTCTCCTGAGTCCAGGCTCAGACTCTCTCATCCTGCTCG  
TTCTCTCTTGGGGAGCCACAGTAACTTTGAGCAACTTTGCATGGGATAGA  
ATGGCCTATTAGGGGCAGCACAAAGACCCCATGGAGGGAAGAGTACAGAA  
AGGGAAAACGATAATCATATTTTTTTAAGATGTGCATTTTCTTAACAAA  
TGCTCTAGTACTTGTCCAGACTTTCAAACCTAAAAACCTAAGCGTCCTTT  
TCTTGAAGATCATCAAAGGCCCCAGTGGTCCTTCAGGTATGTCAAGCTTT  
CTAGAAAATAAAGGTAAGTCATAATCACTTAACACACATGGCTAAATGGC  
CATTTCTTCTAATTTATCAGCAACTGTTACATATTTCTATACTAGAAAA  
AATTTATATTTATACTCAGGGTGGTAAGTTAAATTTGCCATCGAAGTAAA  
GCAGAAAGAGCGTAGCATGTATGTATATGTAACCTCAACTGTGCATGAGAC  
AAAGATGTCTTGAGGAGAATGAGTCTAAGATGCGCCTGAGCAATAGTACC  
C

## &gt;Contig1

GCACCCATGTTTTCTAAAGGGCATAACCAGCCATAATAACAGGATGGGTGAG  
GATATAGACAGCAGATGACAGAGAGGAGAGTGAAAGCTGGGAATCCCAGC  
TAAAGGCATCAGGTTTATGGAATGAGTAGGGGACAACTACTGTGTGTGTTT  
ATACACACATGTATATGTGTGTATATGTATACATGTTTATGTATATATAT  
AATTATATGGTACCATTCTAATTGACAAAATAATCTATCACATTTTACA  
TTATCAGATTTTACATCTATTGTTCTAAATACACTCAGTCATCAGCCCTG  
TGTGTGGGCTCTTACCCATCCCCATGCACACCTCAGCTCAACCACTGATG  
GATGGATCATCTGCCTATCAGAGGTGGCATATTCAAGGTGAATCCATGGCC  
ACAGCTGCAGCACTTCCTACCCACGCAGAAAGGCTCCACAAGAGGAGGCA  
CACCCGCTCTGACTGTCCCTAAGCTCCTGACATCTTACCCCATGAAACT  
GCTGCTCCTGGGTGCTTCCTGCCTTGCCTGCCCTGCCACCCTTGTAAGTCT  
CACCATTGACACAGCTGGTGCCCGATGCAC

## &gt;Contig2

NAAAACGAATCGTCACTATTGAAGCCTGTCTCTCANC GGATCGTGACTAA  
GAACCCCTCCTTGCTTCAAGTTGTCTGCCTTTCTAGGCAGAGCCACCC  
TACATCTTAAATATATTGATTGATGACTTACGTCTCCCTAAAATATATAA  
AACCAAGCTGTGCTCTTACCAACTGGGCACATGTGGTCAAGACCTCCTG  
ATGCTCTTGTGATGAGTGGGTGGTGTCTCAACCTTGGAAAAATAAACT  
TTCTAAATTAAGTGAACCTGGGTGAGATTTTGGGGTTTACAGCAACAA  
TTTAAAAAACTCACCATTGACCTGAAATTTTACCTTATGTGTCTCTCA  
CACTCCTCCATGAAAATAGACGCCATCCTATGAGTTCCCTCAGCCATGTC  
ATGCCACACTTCCAACATGTGTCCCATCCACCATCTGTCTTCTTATTGC  
TGCATCCTACCCAGGCCCTGATCTCTGGACCCATTGTTGTATAATTAAGA  
ATTTGGGGCTGGGCATCGTGGCTGTGGCTCACTCCTGTGATCTCAACATT  
TTGGGAAGGTGTATTAGTCAGGATTCCTCCGAAGGATGCAACCCTAGGGA  
TCCTCTCTATGACCCATATGTCTA

## &gt;Contig3

CGCGCTCAACCGACCGATTTGCGCGAACCTGCCCATGCCCGAGGACAGTG  
TAATCCTAAAACGTCCCCTGAATCATAAGGATATGAGTGCAGAAAGTACGG  
TTCCCTCTGTCAACCACTTTCTAACAACGCTATGTCCGATCCGTGCACTAA  
CCCCGCCCAAGTCACTGAAACACTGATGGGCGCTTCCCTCTACAGGTATCC  
AGGGCCAATACCACTACTCCCTCCTCCCTGTCCCCCTTCCACTCTCTAG  
AGGCCGCGGATGCCATCCTCTATTAGCACAACCGAAAACGACGGTGAAAG  
TACCACGAAGCTCACGATCTGATCGGTGCGCCAATGCGGTTACAACGGCT  
GTCATCCCAACCCCGTCCCATCCTCCATATTGCCCCCCCCCTATGAGGAT  
GGCCCTATCATCATGACCTCCAAAATTCTGTCTATCTCCCGACGTAATGCC  
GCCCCCTCGAACGCTGACACCATCAAGTCNGTCACCTCCCAAAATACTCC  
TCCTAATCACCAGGCCGAGTATCCCCGTTCCACAATACCTCCTTGAGAC  
GGGCCGATATCACACAC

## &gt;Contig4

NGGAGTTTAGGTCAACTAGTAACAAGTGGGATTTGCGACTCAGGTCTATC  
TAATCCTCAAACCCACGTCTGGACCCCTACACAGACTGCCCTCCCTCAG  
TCCTCTGTGTGGCCTCAAGAAGGGTCTGGACATTCAAGTTTAAAAATCCA  
TCCAAAGAATCTATGGACCCAGTGGTCTCTGGAGTCAATGTTCTGAGGCT  
CAGAAGGGCCAGGCAGGAGGGAGCCGCTCTACACAGTCTTGAGCAGAGT  
GGGCTGTGTCCCGGCACAGCAGGGGAGATCATAAACAGAATTCTGCCCTG  
GGCCCTATTTAAGTAGGACCTTTAGGCTGCCGGTGTGATGACCACAGGTC  
CCANGTCTGCACGATTGGCTGTGTGTGGAATCTTCACTCCTTGCGGCC  
TTGTCTTGGCAGAGAGCACCGCTGCTTCTCTGATGGCCACCAGGGGGA  
GGCGCTCCCCTGGGAACGTTTGAANGGGAGCCTCACCCACACGTGCCT  
TCCGTGGTACCCAGCACCAGCTGCTACCCATGGTTACCCACAGGCCAGC  
TCTGCTCTGAAGAAGGAGGAGTGGTGGCGATCANGCCTTGCTGTCATCCC  
GTGGCTGCCCTTTCTTTTCTTT

## &gt;Contig5

GGGAGCTAACCGCTCACTGGGATTACAGGTACGCACCACCGCCTGGCT  
AATTTTGTATTTTATAGTAGAGACGGGTTTCTCCGTGTTGGTAAGGCTGG  
TCTCGAACTCCCAACCTCAGTTGATCTGCCCCGCTCAGCCTCCCAAAGTG  
CTGGGATAACAGGTGTGAGCTACCATGCCTGGGCTTATATGTTTCTAGTC  
CAAACATTTAGCTACCTTTTTTTTTTTTTTGGAGACGAAGTCTCACTCTGT

TGCCCCAAGCTGGAGCACAGTGGCACAAATCGTGGCTCGCTGCAGCCTCAAC  
CTCCTCAGGCTCAGGTGATTCTCCACCTCGGCCTCCCTAGTAGCTGGGA  
CTACAGGTACGCACCACTACACCCTGCTAATTTTTTTGTTTTTGTATTTT  
TTGTACAGATGGGGTTTCTTCATGTTACCCANGCTGGTCTTGAACCTCTG  
GGCTCAAGCAATCTGCCTACTTCAGCCTCCCAAAGTGCTAGGATTACAAG  
CATAAGCCACCATAACCCGGCCTACCTACTTTTAACTTGTGGAATTTTCTA  
TAAGGTCAAGGATGCCTGNGGGAACAAAAGTTTCTCCCTTGGTATATGCA  
AGTAAATCCACATGCTGCCTCCC

>Contig6

AGGACTGTAGCTGTTGTCTAGTCACCAGGCTGGACTGCTTGGCATGATCT  
CAGCTCACTACAACCTCCACCTCCTGGGTTCAAGGGATTCTCCTGCTTCA  
GCCTTCCAAGTAGCTGGGATTACAGGCATGCACTACCATGCCCGGCTAAT  
TTTTGTATTCTTAGTAGAGACGGGGTTTCGCCATGTTGGCCAGGCTGCTCT  
CAAACCTCTGCCCTCAAGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTG  
GGATTACAGGCGTGAGCCCCGGCCACATGTAAAAGTTTATATCTCTGT  
TGTTTTACCTTGTTTTTACCTAGTCTTTAGTGATTGAACTCTTGATTC  
AGTCTTTTGTATTTTAGTGGTACTTCCCAGCTTTGTGTCTCTGTGGAT  
GACATATGAGTCTTGCTTCTTCATGCCAATTTAAGAAGACTGAACGGGAA  
TAGGTCAAAGGCATGGCCATGAGCGATTTCTCTCCAGCTTTTCATGGTGT  
TCAGCTTCAAATCTATTACATATTGGACCTGCAAGCCATCATCTTATCC  
ACAGGCTATCATCATAGGTGAATGTAAATTGGGTTAGGTGGCCAAGCTG  
AACGTGAGATATNTTC

>Contig7

AGCATGTTCTCTAAAGGCCTATCAAAGCTGACATCAAAGGGATAAGTTCC  
AGTTACCCAGCTGAAGGGAAGGAGGGTGTTTCAGATAGAGGAAGGATAAG  
CATGACCTATTCAAGGCCAGTGAAAGAAGCGTGCAACGGCCAAGTCAGGA  
GAACCTGAAATTGTGTCAAAGAGCTTGATGCAAAGAGCCGTGGGAGACT  
ATTGGGGGTTTTAAGCAGGGATATAATATTCAATTCAGCATGCAGTAAAA  
GGTCACTGGCACCTGCCATGGGCCAGGACTCGGGCTCTACATGATTGCGT  
CTGTTTTGGAAATATCACCTGGCTGTGAGATGAAGAACAGGTAGGAGGG  
TCACAAAACCTTGAAGCAGAGAGACTGTTGAGGAAGTAAGCTGTTTTGTG  
TGGACTGTGGCAATCACAGAGGCAGAGGATATAAATGCACAGAGACACAA  
GGCATGTGGGAGGCAGAAGGAATCAAATACAATGAGTGATCAGATGTGGG  
GTTAGAATGGTGAGTGANAAAGACATACTCAAGGTGACACGCCAGGTAT  
CTGGGTGGATGGTAAGACATTCTAGGACTAGAATCGAAGAGGAGGTGGGG  
ATGGACATTCCCTCCGTTTAGAGGGGTTACCAGGAGGATTTGCCGGAAC  
ATGGAGAGGATTAACCAGGAATCCGGTGCCTTTTTCCAAACTGGGTGGA  
GGG

>Contig8

GGTGAATGCTTTGGCACGCTGTGTAGATTTTAGGTGACGGGTGGTGACAA  
TGAGTCCGTGTGAGCGCTGATTTTTTCGGCCTTTAGAGCGAGATTTATA  
CAATAGAATTTGGCATGAGATTGGATTGCTTTTAGTCAGCCTCTTATAGC  
CTAAAGTCTTTGAGTGACTAGATGACATATCATGTAAGTTGCTGATAGGT  
TTCCAGTTTTCCGCTCCTAGGTCTGCATATTGTACTTTTCTCTTACTCG  
ACTTAACAGTACCAACCCAGCTTCTCAACGGATTTATACCATGGCACTT  
TAAAGCCAGCATCACTGACAATGAGCGGTGTGGTGTTACTCGGTAGAATG  
CTCGCAAGGTCCGCTAAAATTGGTCATGAGCTTTCTTTGAACATTGCTCT  
GAAAACGGGAACGCTTTCTCATAAAGAGTAACAGAACGACCGTGTAAGTGC  
GAATGAAGCTCGCCATACCATAAGTCGTTTTGCTCCCGAATATCAGACC  
AGTCAACAAGTGTCAATGGGCTCGTATTGCCCGAACAGATTAAGCTAGCA  
TGCCAACGGGATAAACGAGTCGCTCTTGGTGGAGGG

>Contig9

GGGGTGGGGCGCCTGGTGTTTTCTAAAGAGGATCTCCTGCCAGAAATGGTG  
TGCTGACACTGTTGTCCTCCTTGGTGTTGAACTTTGGTGGGAAGAAAGGT  
TGGAAGGGGAAATTTGATCCTTGGATTTAACCCAGTTTGTACTGATG  
CTCACAAGACTAGGGGAAGGATAAAGGCAGGTGAGTCACTCTAGGATGGC  
TCANTGAGCTCCACAGAGCTGGAACCACAGGCACCAGGAGGGATTAGAG  
CAGGCCTCAGTGACGTCAGCTGAGTGAACCAATGAGCAGGTGATGGGTC  
CAGGCAGAGCCCTGTCCTCTTTAGGCAAAAACCTTGAAACACCGTTCCC  
ATCCTAGCCTGTGTTCCACCAAAGCTGGCCAGTCTCCAGGCCCTGCCTG

AGCCCCAAGGAAGTGGTATGGTGAAACAGAAGGGCCATTCTGTCCAATG  
TGTGAGGAACCTTCATTTTCACTTGTGGAAGCCCTGATGTTCAAAAACC  
TCAATGATATCATTTCATTTTCCCCATCCATTCAATGCCATCCAATGCC  
ATCCGTTCAATGCCCTTCCATTCTCTTCAGGGAAATGAAAATTGTTCA  
GAAATCCTTTCTCTTTTCGAGAAACCAACCAACCAAAACCGCGAAATTCA  
CTAAACTAGCCAAGACACAATCCTGGGTTATTTTCTTTTCCCAAACCTC  
CTGTGTTTAAATTAATTCTACCCTGGTTCTCGGCCCTTACTGCGAAGGTG  
AACTCACCTAACCTCTCCCAAACAGAGAAGAACTTCTCTTGGTAAATG  
GGTTTTAACACTTCTAAAAAACCCCC

>Contig10

GCTATGGTTCTAAAGGTAATGGACTATGGCGTACACAACGTCTCGCTCAT  
CGTCTGCCAGGAGGCTAAGGTATCCACGGACAATCGCTGAGCAACAGTGT  
CGTTGATCCATCTCTGTACGCACTTGTCAACATGGCAGGAGTACGGGAGC  
TGCGAGAATCCTCTCTGTGTATGTCACGGAGCATGCCGTGAGACAACG  
CCACGAACGGCCCTCGGAGANANCTACTCTGCAATGAAGACGTACGATAC  
ACACGTAGGAGTCTTAGCTCACCAGCCGTATCTAGGTATACTGTACTCGC  
GGATACTCACTCGTGCATGCGGCAATAGATCGATACGCAGTCTGTCAGCC  
CATGCTCTCAGTGTGTGACCTTCTGGCGGTAGCGTNGTGGGCGCTATTAC  
TGTGCGCAGCAGGCGCNTCGTACATGTGTGCGGTAGCGATGCCAGGAGCT  
GTAACATAGCAAGTCGCCCCCTACTCCTATCACTATCCCTACGCTGGAC  
CGCACTCGAGATCTGAACGCACGTCTTAACCTGCCAGTACTCGTGAGACC  
TATACTCGCAAGCCTTGCTTAGGAGATCCTGCAGCGCCGGCAAAGAATC  
AGCTATGATCCCCTTGCGATTATCGCACACGCACCATAGAGTATGTGCAT  
ATTAACCTCTGAATGTGCTGCAAGCAGACGGTTGCTCAACATATATATGG  
ATGTGGGGAATCGCCCTGGTCACCGCCACTTGGCGTCAGGAGGCACCAG  
CACGTCTGAGTGTACGCACGTTACTC

>Contig11

GGCCGAATGGTGAATTCATCCGTCGTCTCGAGGGGGTGAAAGACGGGGAG  
TTATGCTGTAATGGCACCGCTCACCTGGGCTTATGAGCAGACCTAACCC  
TCCCANAGTCTGGGATTACAGGCATGAGCCACCGTGCCCGGCCAGTAT  
CTGAATCTCTGTGGCCAGGCAGAAAAGTCTGTGTTACTCGTCTCCTTT  
ATCATTCATGTCCATATTCTCCCATTTGCTAACATTTATGTTTCTGCTCC  
ACTGGATTCTTTGGATTTTCTAGAACATACCCATGCTTTGCATTGCCTT  
GGTCTTTGAATATTTGGTCCACTTTTCTGCAAAGTCCCCTCTCACCTTA  
TCTTCTGGTAAACTTCCAGCCAACACCTCTTTACTAACCAGAGAAACAT  
GGTTCAACTGTGCACAGGCTTGCACAGAACTGTTCTCATATTGTCTTGT  
CATTGTCAATGTGGCAGAGATGCACCTTAGATACCTCTTTGAGAAAGGAC  
TCACTGCCCAGCTGCCTGGCAGGTGATGAGCTGATAGCTCCAGCTATAGA  
CTCCTTTAGGGTCAACCTCTGCTTTCCAGTTGAGATCATATCCTTTGCAG  
GGTGGCCTCCCCAGTGATGACTAAGGCAGTGTTACAATGGCCTAGTCATT  
TCTTCCCAATGCTGGACTCCCAATGAACCATCTGCTCCGGAGCTTCCAC  
TGGGCAGTCAGAGACCTTAGCTAGTCTGCCTCCGAATCAGAAGGCTCTCT  
CTTGCCACTCTGGCC

>Contig12

GCTGTGTCTAAAGATTACGGCTGTAGTTCCTAAGTCCCGCCGCCCTCTAC  
TGTGTCCTCTTAATGGCAGTCATTACCATCTTCTGTCCCTCCCCTTCA  
TTTCTTGGATGGTGACTGTCACTTTGCTGCAACAGAACCCTGTCCCAATC  
CTTGATGGTTCAATACACATAGACATTCTTTTAAACAGGGCGGCCTCT  
CAGGTCTTTAATTTTCTTCCCTCCAATAACCTTGTGATGATCCCCAGCT  
TAGCCACTTAATTGCCAGATCATTACCAGTAACTCCAGCCCCCTCTTAATT  
CTAGTTTCTAATATCCTAATCTGTGACCTCACATTCCAACCTCTTCATT  
TTATCCCCTGAGTCAAAAAATCCTTTGATCCATGCAATCCATTAAGTCAT  
CTACCTTTTACCATTCTTCGCCCCACTAGGGTTCTCATTCTTTATTAC  
CCATATGAAATTCCAAGGCCTGTTGGAATCACTCCCTTGCAGCCACTGTC  
AATACTTCTGCCCCCTTTACTTCATCACCTTATGTGGCAAACACACAGC  
CCTGGTGGAGTGCATCCTTACCCCTGCTCTGTGCCAACAGCCGCACACGC  
ATGGCTGATGGAGGTTGGAAAAATCCACACATGCAGTGGGCCCCTGTATGT  
CCATATACGTATCCAACCTCCAGCCTTGCATATGCCTCAGTGCTGCCTGA  
CAACACATTATATGTTTTCTTTAGTTCCTTCAGTCTCCTGGGTGCCTAGG  
TGAGTATCTCAGACATCCTTCTCTCTGCAAAGCTCCAACACCTCCACG

TCACATTCAACTGATGACTGTGTCTCCTATGTCACTTAGATCACAGAGGC  
ATACATAAACAAATCCCAGCCACTGCCAGCACTCTGCACATCTGCGAGCA  
TGGCACCCCCAATCTAGGCCTTTCTGCTGTCACTTGGGGTGAGCTGATT  
ATACTCGATCCTAGTCATTTCTACTTATGCAC

>Contig13

CTTAAGGCCTCCCTCTAACATTTTAAATTTAAGATTGAAAAAGCAAAGATT  
ATTCTGTTTTGGCTGCGCTATAGTAAAGTAACCCCTATGNCAAATTTTG  
ACACCTTATAGTATTTGACAGGGATAAGTATAAAATTGCTTGATTGATAC  
ATCCACACCCAAATGTATGCTGGGAATGATTTTGTTCACGGCACTCATT  
ACTTAATTTTTTAAACTCTTATTTAAATTTGCAATGTTTTAAATGACCAT  
CACTTAAAGTAGTAATCAACAGAGGTTAGGAGAACATAACAATACTCTTT  
CTCTTAGAAAATACAACAGAAATATAATTTTTTACAGTTTTGCTCCCAA  
CTTTTCTCTGTAATAACATGCCTTACTCACCTTTACAATAGGTTTGTGT  
GAGAATCTTGTAATGTAAACCCTGGGTGTTCTGTGAAGCATTTTTAAACT  
TCTAGTTTACACTGACTCTTATTCAAGTGTTTTTAAAAATATATTTAAAA  
AACTGGCCAGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG  
CCAAGGCGGGCAGATCACAAGGTCAGGAGTTTGAGACCAGCCTAGCCAAC  
ATAGTAAAACCTCGTCTCTACTAAAAATACAAAATTAGCTGGGCGTGGT  
GGCGGGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGAAGATCG  
CTTGAACCCGGGAGGCAGAGGTTGTGGTGAACCAAGTTTGGCCAATGCA  
CTGCCAGCCTCTGCAGNGACAGCC

>Contig14

GGGGGCGGGCCGAGTGATCCTAAAGCCCCGCTCGCTTCAACAAAGCCTA  
ACAGTCCAATCACTTAATGCTGCATTTATTCCTGGGGAAGCAAGTCTCCT  
TTGCACTTTACACAGTGAGATAATCAGTTTCTCATGTGGACCACTGGGCC  
AGGAGGGCCTGACAAAGGGCAGTCTACATTTCAGACTGGAACTGCTCCC  
AGAACTATTTCTTTCTAGTTCCCACCTCGGTCTGAGGTGCCTGAGGAGAG  
GGACTCAACAGAGGAAGCAGGAGCATAGCTCAAAGTCTCAGAACATGGAA  
GAGGAAAAGAATCCTCACAAGATTACGTAACCTACAGGCGTGTGCTGCT  
TCAGTAGAAGTTTCATCTCCCTCAATCCTGTACACTTTTCCATACATTAC  
ATACTCAAACCTGGTCAGCCCTATGGAGCAATAGCAGCAAAGTTATTCTTA  
ACAGTAATTAACAATATAAAAGATCCCATTTAAAAATGGTTACTGGTCAG  
CCGGGCGTGGTMNNTCNANCCTNTAACCCCANCACTTTGGAAAGCATGCG  
GGCGATCCCAAGTCTGATATCGAAACATCTGCCTAACATGTGCAACCCCT  
CTCTACAAAATACAAAAATATCCGGGCTTGTGTTGGCGCCGTTATCTCA  
CTACCCGGAGCTAAGTAAGAAATGCTTTACCTGGAAGCGATTTTTTTACT  
TATATCCCCTCTCTTACCCGGGCGGACCAAATTCTTTAGTATAGGAAAG  
TTTATTGTTTTATGCCTTTGTCAAGGCTCTACTGTATCTTTTCTGTCCAC  
TCAC

>Contig15

GGTTCTGAACAACAGCAGGCGATTCTAGCCCTGTACCCGGGGCATTGTC  
CAACACTCGACAGGGCTGAATTCGTCCATAACGGTGTGCCCCCTCTGGGAT  
ATAGGATGAAATGAATTGATCTGAGTACCTGGGATGTAAAGTTACTAAAA  
CGCCAGCTAGGTTACGCCCCGATGCTTAAATATGATCGTGGCCTACACC  
TCGTCCAGCAGAAAAAGTACCCTTTCTTCAACACCACCTCACGATCCTCC  
AATTTAGGAGCTATAAACTCATGACTCTTTATTTACCCCTGCAGATTCT  
TCAATCCAATAGTGTGTGTCTCCCTGTGAACCTCACGGATATACCGATTTT  
CCCCACGTCATTTCCACACGTCGCAATCGCTTAGTCATCCCTATGTATGA  
GAATCATGGATGACTATGTTGAAGTCCATCTATAAAGTTCAACCCCATC  
TCCGTCCCTGATTCCCCCTCCCCAAGATCACCAACGCGACTCGACATATT  
GTTATCGCCCAAGGGACCTCTTGATCCCCCATATCCACTGGTCACTCC  
CCTCTTGGCTGGAAGTCACCGGAAGTTCTCCACATGTTGT

>Contig16

TGCGAGCGATGTTCTTAACTTTAGCGCCATTGACTCGAGCATGGTCATG  
GCTGTTTCCTG

>Contig17

AGGGTGTTCCTAAAGGATACTACGTTCCCTAAAGTCCAGAGAAAAAAAAA  
AAAGTAACATAATGTGGCTTATTTGGTATAAAAAATTTTACAGGAAGCATT  
GTCAAATATGAAATAGTGTGTTTGGTTTTGTTTGGGCTGTATTTGTATAAAT  
ATGTTATTGGTATGTGTTCCAAAATTATAGGAACTCCTATAATTCTGAT

ATGACTTGGTGTACATTATCAGTAATAATTATAATTGTTATGGTAAATTA  
TTGTGTGCCATGGAGGTAACAAATTTCCCTCATCAAGTGTGTCTTTGACTA  
TGGTTGCCCTAAAACCTTTTGGCATTACAGACAATTGTCTTGCTTTGGT  
CCTCTTTAGAAGGTGGTTTTATAATCAGCTATAAACTCTAACGGGTGCT  
CTTGAATGCAGGCTTAAGATAGCTTTGGAGACTGTGACATCAGAATAGAG  
GAAAACTTTCAGTATTCATGGAGTGTGAAATATTCATGAATATCAAGC  
AAAACAGGAATTAACCTTCATAGATGGAACATAAAGAATGCTGAAGTAATC  
TTTTTGACTTTTTTCTTAAATGTTGATCCTTCGTTTTGTTTTTCAGAG  
TCAAGGAAATTTTTCTGTTGAGATATTGACAGCTTTTAACAATTAAGTAT  
ACTCCAGTGAACACAATTTGGAGCATATTTGTGTCTCTCTATATATATTT  
GGAAACAATNTTTGAGTATTCTTAACTTATTGCAATATT

>Contig18

GGTTGTCTGCTATACCAGTAATGGGATTGCTGGGTCAAATGGTATTTCTG  
GTTCCAGATCCTTGAGGAATTGCCACACTGTCTTCCACAATGGTTGAACT  
AACTGACACTCCCAACACAGTGTAAGCATTCCTATTTCTCCACATCC  
TCTCCAGCATCTGTTGTTTCTGACTTTTTAATAATCGCCATTCTAACTG  
GCATGAGATGGTATCTCATTTGTGGTTTTCAATTTGCATTTCTCTAATGACC  
AGTGATGATGAGCTTTTTTTCATGTTTGTGGCCACATAAATGTCTTCTT  
CTGAGATGTGTCTGTTTCTATCTTTTGGCCACTTTTTGATGGGTTTTTTT  
TTCTTGCAAATTTGTTTAAATTCCTTGATGATTCTGGATATTAGCCCTTT  
GTCAGATGGATAGATTGAAAAAATTTTCTCCTATTCTGTAGGTTGCCTGT  
TCACTCTGACAATAGTTTCTTTTGTGCTGTCAGAAGCTTTTCAGTTTAATT  
AGATCCCATTTGTCAATTGGCTTTTTGTTGCAATTGCTTTTGGTGTCTAA  
TCATGAAGTCTTTGCTCATGCCTATGTCTGAATGGTATTGCCTAGGTTT  
TCTTCTATGGTTTTTATGGTTTTAGGTCTTATGTTTAAATCCTTCTTTTT  
TTTTTTTTTTTTTTTTTGAGATGGAGTCTTAGTCTGTTGCCAGGCTGGA  
GAGCGAGTGGCGTGTCTNTAGGACGC

>Contig19

GCATGTTGTCTAAAGGTTTGTCTTCCCTCCAAAATTCATATGTTAAACCT  
AGCCCCAAATGTGATAATATTTGGAGGAAGGCTCTTTGGGAGGCAGAGCC  
CTCATGAATGGGATTAGTAGCCTTATAAAAGAGACCCCTGAGGGCTCCCT  
TGTCCTCCACCGTGTAAGGATGCAACAAGAAAGTATGGTCTATGATCC  
AAAAAGCAGACCCCTTGCCAGGTACCCAATATGCTGGCACTTGAACTTCCC  
AGCCTCCAGAACTGTGAGAAATAAATTTCTATTTTTTCATAAGCCACCGAG  
TCTATGGTATTTTGTATAGGAGCACAAACAGACTGATGTGCCACCCAAC  
CATGATTATACGTGTAATTTATGGTTTCTCTGCTAGTAGGGATGCACCAT  
GGGGTTAGGAACACGCTTTTCTTATTTCCACACAGTCTTAGCTCTAA  
GCATGTTCTGAATCAAAGATCCCCATCTTTTATGAATGAAGGAGTCAGT  
GAATGAATTAATGAAAGAACTGATAACCCTCAATAATTATTCCAGCCTTT  
TATACCTACTATTAA

>Contig20

ACGGTTCTCTAAAGACTTTCAAGAGCTGGATTTTATGCTTTAGGTGAAGG  
TGATAAAGTAAAGTGCTTTCACTGTGGAGGGGGGCTAACTGATTGGAAGC  
CCAGCGAAGACCCCTTGGGAACAACATGATAAATGGCATCCAGGGTGTA  
TATCTGTTAGAACAGAAGACACGAAAATATATAACAATATTCATTTATC  
CCATTCACTTGAGGAGTGTCTGGTAAGAACTGCTGAAAAAACGCCATCAC  
TAAGTAGAAAAATTGATACCATCTTCCATAATCCTATGGTACAAGAAGCT  
ATATGAATGGGGTTCAGTTTCAAAGACATTAAGAAAATAATGGAGGAAAA  
AATTCAGACTCTGGGAGCACTGTAAATCACTTGAGGTTCTGATTGCAG  
ATCCAGTGAAGGCTCAGAAAGACAGTACACAAGACGAATCAAGTCAGACT  
TCATTGCAGAAAGAGATTAGTACTGAAGAGCAGCTAAGACACCTGCAAGA  
GGAGAAGCTTTGCAAAATCTGTATGGATAGAAATATTGCTGTGTTTTTA  
TTCCTTGTGGACATCCAGTCACTCGTAAACAATGTGCTGAAGTGGTTGAC  
AAATGTCTCAAGTGGTACGCAGTCATTACTTTCAAGCAAAAAAATTTTAT  
GTCTTAATCTAACGCTATAGTAGGCATATTATGTTTCGTATTATCCTGATT  
GAATGTGTGATGTGAAGTGAAGTAAAGTAATCAGGATTGAATTCATTAG  
CATTTGGTACCAAGTAGGAAAAAAAATGTAAAGCCAGTGCTTAGACACA  
GC

>Contig21

CGCTGTCTTAAGAACTGGGCTAGGAGTGAGCAGTGAGCCAAGATCGCACC

ATTGCACTTCAGCCTGGGCAACAAGAGCAAACTCCATCTCAAAAAAATA  
CATATATATATATGACCCATAAAAAGGAGATAAATCAACACTTCAGAACT  
GACCCAAACTTGCAAAGATACTATAATTAACAGAAAAGGACAGTTTACTA  
AGTACTCCGTATGTTCAACAAGTGAAAGATTAAACATATTAAGTAGAGAT  
GTAGAAGATATAAGAAGATCCAAAATGAACTTTTAGAGTTGAAAACTACA  
ATATTTAAGATAAAAAATACACTAGGTGGGATTAAAAGTAGATTACACATT  
GCATAAGATAAAAAAATGAGCCTGAATACAGCACAGTATAAACTATCT  
TAAACAAAACACAGAGAGAAAAAATAACTTTAGAGACTTAGCTCTTATC  
CTCTATTTGTTTCTAAACAGAGGATAAGGGGCAGAAAAAATGTTTGAAGA  
AATCATGATTTTTAAATTTCCAACAGATAGGAATAGCACTGGGTAGTC  
ACAGGAGGCTGGAAAGACCCAAACAGCAGTTAAAACAGGAACTAGGCCAA  
GAAACCAAGGATAACAGTAAACCTAACTAAGGGAGAGAAAACTGACAA  
AAGCTGACTTAGGATAACTGAC

>Contig22

CCTGAATATAAGCCGCAAGTAACCAATTAATTTGTTTTCCAAAATTGTA  
TTAACAATCTATGAAATTTTTATCTTGACCATAGCTATAACTTCCAGAAG  
CCTTTTATAACCTCTATAACCTTTATTAAGGAGTAGGTTAATGCTTCAAG  
AAAACCTTGTTAATCTGACACAGGACCCATATGCTGATCTTGTCATCAGTG  
TGGCTTGGACATCAATGATTATGATTAATTTATAGAGAAATTGAACCTAT  
TTTATCTCTCAAAATTGGCCCTTACAATCTCACACACCCACCTCTCCAC  
TATAGTTCCTGGGCCTTGAGTTGAATAGCTTTAATTTCTGGCTCTGTGTT  
TCAAGAATGCAGTTTATTTTGATTGGCATTCTTACCAGTCCTGAAGATG  
AACCTTTAATTGCTGTCAGTATTTAAGATTTAGCAGGACTTGTCTTTTA  
AGAACCAGGAGTCAAGCCCTATAACTCAATGTCACAAGGACTTTAAAAGC  
ACATACATAAAGATATATGGATGTAATAATCATAATTTTTAAAAAATTGT  
ATTAATCTCAGTGTTTTCTAAGCAAACCAAACTTAATAATAATGGCATA  
GAAATTATTTCAATAAAACATAAAATCTGTTAAGCCAGTTACCAAAAAGGC  
AAAAGAAAAGACCTTCTGCAATGCACAGAATATTATGTTGGAAGAAAAACA  
TTTCCTTTAGACCTTTAAGAAAACATTGTTAGCATCAGGACACAACAAC  
AGAATCTGAGGGTAAAAAACGTATATGAGCTGAAGGGAGTTGAAGGAGGG  
CATTACTATTTCCACCTTTTAAAGGGGAGAGAAAACCTAAAACAGCAA  
GATGCAATAAAAAGCTGAACTTTGGGTTAAAAAAAATTCTTAAGTCTCTT  
ATAATTTTATTAAGAGTGAATCAACCCCGTAAGAAAATTTCAATTGTTCTAA  
CCAATTTTTTAAATATATAAGTAGTTTTTTAACATCAACCCAATCTCTAGA  
AAGACCATTATAATTTCCCTTTAATTATAGACAACCTTTATCATATAAAAG  
TTTTTTTTAAATAAATCCTCTTATTGTGACTTACACAGACTATTCATGACA  
TGCTTGGACTTTCTGGTTTGTCTGTAACATCCTTTTCTTTCTTTCTTTCT  
TTTTTAAATTTTACTTTACGTTCTGGGATACATGTGAAGAACATGGAGGT  
TTATTACGTAGGTGTACATGTGCCATGGTGGTTTGCTGCACCCATTAAAC  
CGTCATCTATATTAGGTATTTTTCTAATGTTATCCCTCCCTTGCCCCC  
CACCTCCTGACAGGCCCTGGTGTGGGACATCCCTCCCTGTGTCCATGTG  
TTCTCAATGTTCACTCCCACTTATGATTGAGAACTGCAGTGTTTGGTTTT  
CTGTTC

>Contig23

GCTAAATATAAGCTATGATAAAACAGTTGGCCCTCTGTATCATGGGTTTC  
ACAACCTGTGGATTCAACTAACTGTGGATGAAAAATACTTGGGAAAAAAG  
AATGGCTGCATCTGTACTGCACAAGTGCGTGCTTTTATTCTCGTCATTAT  
TCCCTAAGCAATACAATATAACAACATTTATATAGCATTACGCTGTAT  
TAGGTATTATAAGTAATCTAGAGATGATTTGAAGTATACAGGAGGATGTG  
CTTAGGTTACATGCAAAATATTATGCCACTTTATATAAGGCCCTTGAGCCT  
CCTCAGATTTTGGTATCCATGGCAGTCCTGGAGTCAATTCTCCTGCAACA  
TCTCCATTGTTTCAAGATTCTCTCTATATCATGTTTATATCAGAAAATCT  
ACATAAGATTTTTTAAATGTGTTTATATAGGTTTTGTGTATTTTTGGTTGT  
TAATCCCTAGATATATGCAGTATTTATTGCTATTATGAGTAGTGTCTTCT  
TACCATGTATTCTAGTTGGTTATTGCTGACAGAGAAATGTTGCTGGTGT  
TCTAAGTTACCTTGTTTCTAACAACCTTGCTGAACTCTTATTAGTTCTCA  
TAGTTTTTAAATTAATCTTTCTTAGTTCTGATAACATAATCTGCAATAAT  
GACAATTTTATATCTTTCTTCCAATGCTTATATCTCTCAGTCCCTCTTA  
TCCCAAAGTATTTTCCAGGATCTCCACTATAACATTAAATAGTAATAAGA  
ATTTCTGTCTTGTTACTGATCTTAAGGAGAATAAATTTAAATTTCTCTG

TCAGGTTTTATGCTTGATATAGATTTGTGATATATAGCCTTTCACAGGT  
AAAAAAAAATGCTTTCCTAGTAGTCCTAATTTTTTAAAAAATCATCATA  
AATAGATGTTGAACATTATCAAATGCTTTTTCTGCATCTATAGAGATAAT  
CATATGGTTTTTTACTATTTATTAATGTAATGAATTAGACCAATTTTCTA  
ATGCCAACTCTTTCTTGATTTGTAGGGTAAATCCTATGGGATCATAAAA  
TACTTTTAATACATTGTTAGATTTGAAGAGTTAACGCCCTTATTTAGAACG  
TTTTTCAGTCACATCCATAAGTGAAATGGCACTATAGTGTCTATTACTATT  
ATATTTTTCTGGTCTGAAACCAAATTTATACTCACCTCATACAGTAAGT  
TGGGCAACTTTTGTCTTTTTTCTGAAACAATTTGTGTATAGAAGAAAT  
TAACGTGTTCTTGAAAGTTTGATAATAATCATCCAGAAAATTATCCCAT  
CTAGGGCTTTTACAAAAAGGAGACTCTAGAATGCCATTTTCGGTTTCCTTG  
ATGTGTATTGGCCTCTTTCATTTAGGCTTTTGGATTTTTTAGGGCATT  
TTCACTATAGGCTTTTTACCGG

>Contig24

CATAAACTTCAGGTTGGATGTTCCGGTCAAAGTGGTCCGGCGATGCGAAAA  
CGAGAGGGCTCGAGGACTGGGCAGAGAACTATTTGAAGGTATCTCTCAGG  
GGAAACCAAGCGGAAGGCGGGGAGTAAATTTGGGAGGGAGCGACGGCCTT  
CAAAGAAGGGGCTTGCAATTAGATCGGCGAGATCCGGGAGGGTCTGGTGGG  
GAGAAATGACTAGAGGACAAATCTAATGGAGAGACAGACGGAGATAGATA  
TCGTGACAGAGAGAGGGACAGTGACAGCCACAACAGTGCAGGGTCCATG  
AGTACAAGGCCCTTAAGTGTACCCCCAGCCGGAGTCATGGCAATTCGAT  
TCCTGTACTGACCACCCAGGATTTGGGTAGACTGTACGAGTTAATGAGCA  
TGGTCCCCAACAAGACTGCTTCGACCTCAGATGCAAAGCACACTTCAGGG  
GTCCCCAAGCCACTCATGTTTTTTGAATGACTGCCATAAGTTCAAAATT  
CCCACAATTCTCTCAGATTCAATAACTGGGTATAACCACTCATAGAATC  
AAGAAAATGCTATCATTATTATTACAATTTTATTATAAAGGATACAAATC  
AGAAGGACTAGCCAAATGAGGAGACACATAGAGAGAGGACTAGTAAAAAA  
CAGAGCTTCTGCGTCTTACCTTCAAGGAATCAGGATGCACCACCCTCCA  
GCACATCAAGTGCTCATCAACCAGGAAGTTCCTCTGAGCTCCAATGTCCA  
GAGATTTTAGGGGAGGATTCAATACATAGGTATCATTGATTAAATCATTGG  
CCATGTACTTGAACCTCAATCTCCAGTGTCCCTCTTCTCCCTAGAGGTCTG  
AAGGGTTGGCTAATATCATGTGGCTCAAAGCCCCAACTCTAATTACCTTT  
TTGGTCTTTTCAGGGACTAGACCCCATCCTGAAGCTATCTACAGGCCCTG  
CCATGAGTTAGCTCATTAAACATAACAAAGACACTTATATTACTCAGAAAA  
TTCCAACAGTTTTAGAAGCTCCATGTCAGGAACCTGGGACATAGATCAAA  
TTCTTTTTTTTTTTTTTTTTTTTGGAGACAGGGTCTTGCTGTGTGCCCAG  
GCTAGAGTGCAACGACAGATCACAGCTCAATGCAGCTTCAACTTCCCAGG  
CTTAAGTGACCTTTCCACCTTAACCTTCCAAGTATCTGGGACCACAGAAA  
ATGGCTAATTATCCTGGCTGATTTTTTAACTTTTTTTTTTTGTAGGGATG  
GGATCGCCCTGTGTTGCCAAGGTTGGTCTCAAACCTCCTGGGTTCAAGCAA  
TCATTCTGCCCTGGCCTCTGTGATGGTTAATACTGAGTGTCAACTTGATT  
GGATTGAAGGATACAAAATAATTTTTTTGGGTGTGTCTGTGAAGGTTTCG  
CCAAAAGACATTACTTTGAGTCAGTGGACGGGGAAATCCCCCTTCCCCA  
TGGGACGGGGAGACCCCCCTCCATCCAGGTAAAAAATCTAATCACCTGC  
AATGTGGCAGAAATAAAGGAGGGGAAAAACGGGGACCCCTANATGGGTTA  
TTCTCCACCTAATTCTTCCCCCAGG

>Contig25

CCATGATTTTCAATTTCTACAGACCCTGAGATGAATTTGTCAATTGCCACGG  
GGTCTGAAGTTCAAATACTCTATTTGGTATCCTGCCCTGTGGTTAACT  
GTGATCAATTTCACTCACCTTGTTTATGATGAGAGGTGCCACCATCTGGCC  
TCCTCCACTCTGCAATCCTGTTAATTCCTATCAAAGCTGAAAACCTGCTG  
CAGCACCCACACCATCACCTCCAGCCTAGAGAGGGAAGCTACCAGTGAGC  
TCTCCTGGATGCCGGTGTGCCCTCGCCAATACATTTCTTCTTAGTCCCT  
TGGTCATCCTGAGGTGTGTGATTAATGGACAGCTATGTGGATTGCACATA  
ATAGATGTACTCCAGCATCTTCATCCCTGATTTTCTTTACAGAAATCAC  
TCAACCTTAGCAACATGTGAAAATCACCTAAGGACATTCTTTAAATCCCT  
CTGTCCACATGGCAACACAAACCACTTAAATAAGAATCTCCAGGGAGTCA  
CTCAAGCATCAATGTTTTTTTAAAGCTCCAATTTTAAAGGATCATTACATTA  
TGTCGAAGAAATTATAGTATTTTACGCCCTTACTGACTGTAAACCACCACCA  
TATCTAAGCATCCATTAGTCAACCTAGCAGACAATAAACTAACATTACCT

CCAGGTA CTCAAATCAATTCATTGCATCCCAAATCCCAGATGGGCCCACC  
CTTATTGACAAATTCAGCCCAATCTTGGTTGAACACATTTAGAATATATT  
TCCATGAACAATATCCGGTTGACGAGTTTCTTTAACTTTTTGGAGTTTAA  
GCCATTTCTTTTACAGTAGCCTTGTTAATTCCTGTCAATGCTCCATGG  
GGGT CATGAAGAGACCTCTTATTAAGTGTGAAGCAACTTGGCTCAGGTGC  
AGACACTCAAATGCTTCACATGCAGTGGGAAAAGAGAGTGATTGTCTAC

>Contig26

TTTAAAAAGAACTGAGTCTTTATTTCAGTCGATTCTTCTAATCTATGAACA  
TAGCATCTCTCTCAAAGCATTTAGTCCTTCTTTAATTTCTGTCAATTAATT  
TTTTAAATTTTTCATCTCTAAAGATTCTGTATATGTTTTGTTGAATTTATG  
CTTAAGCATTTTCACTTTCTTGGTAACAATTATAAATGATTTTGTGTTTTT  
TATTCCACTAGTTCATTTTTCAGTGTGTAGAAAAGCAATGAATTTTTGTGT  
GTTGATCTTTGTTCCAACATCTTGCAACATTATTGAACTCATTTATTAGT  
TCTAGGAGGTTTTTTTCATTTTTCTTGTAGATACCTTGAGATTTTCTATAT  
AGACAGTCATGTTGTCTGCAAACAGGCACAGTTTTATTTCTTCTTTTCA  
ATCTATATGCCTTTTTTTTTTTTTTTTTTGCCTTATTGCAGTGGGTAGAACTT  
CTAGCACTATGTCAAATAGCATTGGTGAAAGCAGACATCCTTGTTCTTTG  
TCTTAGAGGAACATTTGGTCTTTAATCTTGATTTAAAAAATTCCTTGCAC  
TAAGTTACCGTGTTTTGCGGGAGGGAGAGGTGGGGTGAGGTGGGGATTTC  
CCCTAATGTTTACAAGCTGGGATTTTCTTTTCTGTGTCTAATTATTTT  
CCTCATTGGCTTTGAAAATCTGATAAAACATTTTAGGACTGTGTATAAAA  
TAGAATTAGCCAAGTGCAATGTCTTTATTTCAGAAGAAATTTTCATGGACGT  
TGTGCCTACTCTCTTGGCTTCTTGGCTTCATGGCTTTCAGATCCACAG  
TAAGCTCTGGATAGTAGAAGTTATAGTAAGACTGACTTCTAAATAAATGA  
AGTGACTTTTAACCTTACTGATATGGCTTAAAGAAAAGGAGTGGCCTTTAA  
GATCCATGAACCTTCTCAAACAAAAGTGATAACGTTATCTCCATGCATATA  
TAATACTAAATATAATGCAACTGAGAGAAGTAGGCTGTGGTAAGAAAGGA  
GACCCAAGTGCCATCTGAAGGCAGCACTTACCACTCTGCTTCATCCCACC  
GAGGAAACAAAGCATGAGTATTGCCAGATTTTCTTCTGTTTCAAGAAAAG  
CCAGAAATCCAGGTTTTTGGCGTGAATGTCTGATTTTAAATGTTGGGAAC  
TAATTTATATTTTGAATAACATTTGTGTGGGACAAGTGAACCTGTATGTG  
GAACTGCTTTCTCCAGTGGCGACCAGTTTGGACCGTTGATACTCAGCAA  
GTTTCAGCCAAGTGCGCCTTGTCATTGTCTAGTCATCAAGGTGATGTGTGAT  
TGGTCAAGCAATTAATTTTGCTCAGCATCTCGTGTGTTTTCAAAGAAGT  
GAAGGTTCAATTTGC

>Contig27

TTTCAGAGCACAATGCGTATTCATAGTATATTGACTTAATTTCTAAGTGT  
AAGTGAATTAATCATCTGAATTTTTTATTTTCAGATAGGCTTAACAAATA  
GAACATTCTGTATATAAATGTGTAAATTAGAGTTAATCTTTCCAATCACA  
TAATTCGTTTTATGTGAAAAAGGAATGAACGTGTTCCATGCTGGTGGAAAG  
ATAGAGATTATTTTTAGAGGTTTGTCTGTTGTGTTTTGGGATTCTGTTTTT  
TTTTAAATTTGTAATATGTACTTGTGTGAATGATTTTTTAAATGATTT  
TACCATTTTTGGAAGGGTATTTAATGATAGAATATCATCGAGCCAACATG  
CACTGACATAGAAAGATGTCAAAGATATATTAAGTGTAAATGCAAGAGG  
GAAAACACTATGTACAGTCTGAGCCAAATCAAAGCATGTATGTTTTTAT  
ATGTGTACAACAAAAGGTTTTGAAAGATATGCGCCGAATTGTTAAATGTG  
GTTTCACTTGAGGGGGTGGGAGGATGGGGCCCAGAGGGGTTTTTATGGG  
GGCCTTCACTTGGTATTTTTTTTCATTTTGTTCTGTTTGAATTTTGT  
TTTCTTTTAAATGGAGTTTCACTCTTGTGCGCTAGGCTGCAATGTAGTG  
GCGTGAACCTCAGCTCACTGCAACCTCCGCTCCAGGTTCAAGTGATTCT  
CCTGCCTCAGCCTCCCATGCCTCCTGTGTAGCTGGGATTACAGGCACCCA  
TCACCATGCCTGGCTAATTTTTGTATTTTCAGTAGAGATGGGGTTTCACC  
ATGTTGGCCAGGCTGGTCTGTAATTCCTGACCTCAAGTGATCCACCCACC  
TTGGCCTCCCAAAGTGCTGGGATTTTCAGGTGTGAGCCACCACGCCAGCC  
CTGTTTAAATTTTTTATAAGTATGTACTACTTTTGTAAATCAGAATTATTA  
GAAAGCATTTTACTGATTTTAAAGCTTAGACATGTTCAAATGCCTGCAA  
ACTACTTAACACTCAGCTTTAGTTTTTCTAATCCAAAAGGCCGGGCAGT  
TAATCTTTTTTGGTGCCAATGTGAAATTTAAACGGTTTTTATGTTTTCTG  
TGTGTGAATGAAAAATATTTCTGAGTGGTGGTTTTTTGACAGGTAGACC  
ATGCTTGTCTTGTTCAAAATAAGTATTTCTGATTTTGTAAATGAAAT

ATACAATATGTCACAGATCTTCCAATTAAGTAGTAAGGGTTTATCCTTAA  
TCCTTGCTAATTTAAGCTTGCATAAGTCACTTTACTAAAAGATCTTTGTT  
AAGCTAGTATTTTAAACATCTGTGAGCTTATGTAGGTAAAAGTAGAAGCA  
TGTTTGTACACTGTTGTAGTTATAGTGACAGCTTTCATGTTGAGGTTCT  
CATATCACCTTGATCTTGAAGTTTCATGTGAGTTTTTACCATTAGGATG  
ATTAAGATGTATATAGGACAAAATATTAAGTCTTTCCTTTACCTAAGTTT  
GCTTCTTGACTAGTAATAGTAGTAGATATTTCTGTAATAAATGTTCTCT  
CAAGATCCTTAAATCTCTTGGAATTATAAAATTATTGGAAAGACAAGA  
ACAGTTTTTATTATTATATGCATTATTATCG

>Contig28

CTTCTCAAGAAAAGGGAAGTGGAGCAATTAACATATGTAATTTTTTTT  
TAAAAAACCTAAACCTAAACATCTACCTATATACAAAAATTAATTAACA  
ATGGATCATGGACTCCAATGTAAACATGAAACTCTAAACTTCTAGAAAA  
AAACTGGAGAAAACCTTTGGTACCTATGACAAGGCACAGTTTTTAGACT  
TAACACTAGAAGTGTGAACATATACAAGAAAAATTAATAATTGAACCTT  
ATGAAAATCAAATTATTTGCTCTCCAAAGACCCTGTTAAGAGGATGAAA  
ACTAAATTACAGATTGAGAGAAAATATTTGTAAATCACATTTTGACAAT  
GGACTTGTATCTAAAATATCTAAAGAACTCTCAAACTCAACATTAAAAA  
AAATATCTAATTAGAAAATGAGTGAACATTTTACGAAAGGGGCCTTAG  
ATTAGCAAATAAAACACTTGAAAAGATACTCAGCATCACTAGCCATTAGA  
AAAATGCATATTAACCAACAATAATGTATCGCTACACACATATAAGAAT  
GGTTTATGAAAAAATAGTGATGACACCAACTGTTAGTGAAGATGTGGAGA  
AACACTCATACTTGCTGGTAGAAATGTAAATGGCATAGCCACTGTGGA  
AAATTATTTGGCAGTTCCTTTTAAACTAAAAATCAATCTACCACACAAC  
CCAGCAATTTTATTACAGGGCATATATCCAGAGAAATGAAGATTTATGA  
TCACACAAAAATCTGTACACAAATGTTTTATGGTCACTTTATTTCATAATA  
GCCAAAACCTGGAACTATCCAAATGTCCTTCAATGGGCAAAGGATTAAA  
CACACTGTGATACATCCATACCATGGAATACTACTCAGCAATAATAAGGA  
AAGAATTACTGCTACACACAAGTTGGATTAAACTCAAGGAAATTGTGCTG  
AGTGAAAAATTAACAAGCCAATCTCAAAGGACACATACTTCATGATTCCA  
TTTGTATAACATTAACACAATTAATTACAGAGATGGAGAACAGAAT  
AGTGGTTGCCAGGGATTATACATGGTGGACGCGGTGAGGCGGGCCTCCAC  
GCCTTGGAGATGAAGGGGGCTACACCCTTTAAAGCACACCCACGAGAGAG  
TTTTGTGCGGAGGGGGCCCAATTTAAGTACTCCGCCCCGGGGGGGAACAC  
AGGGGCAAACAAAAAAATTTGGCCTTGGGGGTGACCAAACACACAAAAAA  
AAAACAAACACACAAAAAAACAACNATGGGTGGGAGGATTAATCGCCAAA  
TCTGAGTAAGCTATCTGGACAGTACCAATATCGATTTCCAGTTTTGATG  
TTGTACTATAATAATGCAAGATGTTAACATTGGAAGAAGCTGGCTGAAGG  
GGGCTCAGGAACCTCTCTGGACATTTCTTTGTACCTTCCTGTGAATCCATC  
ATTATTACAAAATAGGACATTTTCTAAAGGTTAAATCATTTTTAATTTTAA  
AATGTCCTGTTACTGTTGAAACTCACATCTCCATATACTGATCAAGAAC  
AGCACTAATGGCCCTGGCCTCCAGGAATTCACAATTCCTACTGACTTTT  
CTTTGAAACCTTGGCCAAGTCGCTTCTCTCTCTGGTCTCAATTTTTCA  
TCTTCAAAATGAAGATTGAATGACTATTAATAATCTCTTGCAATTCCTGAG  
ATGAAGGGTCTTAAAGGAAGTGAAGAGGATGCCATGTAATGTAAATATGG  
GTTTTTACTCCATCAGCCAGCCAAGACAGAGGGCAGACACCAAGACATGG  
TAACCAAGGAGGCCATGTGTAAACAAAGACCATTTAGACTTATGCTCTGG  
CCTTTGCAGCCCAACTGGTGTGGCCAGTTGGTGGGGTATGAAGAAAATGG  
GGCCTTCCAGGAACCATGTTGAGTGGAGATAAGCAGGGAGGAATGCAGAA  
GACATGGGGGCAGTGCCAGTCTCAGCCCGAGCCAGCTACACCCACACATG  
GTTATGAAGACTGACAGCCTGTAAGNTGAACACAGCCCTGCCTCTCTTA  
GATAGGC

>Contig29

GCAATATGATCTCAGATGTGGATTTACTGTAAAGTTCATCAAATTTAAA  
TTTCAGAACACTTAATCTGCAAGAGTCCTTTCCAAGACCTTATACCTAAT  
TTTGTGTTTACAATTTTATATTTGTTTTCTTAAAGAAGACCACCAATATA  
AACTATATCCAGCCTTCATGATAAGTACATAGGAACTATGCAAATAAGG  
GGGAAAAAAACAAAGAAAAATACCTAGTTTACTAATGGTTCACCTCTGA  
ATAGCACATATTATAATGATACAAGCACTCATTACTAGTCTAGGAAAT  
GAAGATATAATTGCATTAGGAAGATCAAGAGGTAGGAAATGTGGATGTGT

GTGGTATAGACTAGGGCAGGACAAAGAACCTAAATCCTCATTTTCTAAAG  
ATAATTGTTAATACGTAAAACTCAAAATTCAAGAAGTAACAGTAAAGCG  
GTCATTAAGAAACAAGCACTAAACACCAGATAGGAAGCGAGAGATGGGGG  
AAGAGGGCGACAATCTGATTATTTTTTGAACAAATTTTGTAACCATT  
TGACTGTTTACATGTAGAACTTGGATCTTTTTTAAAAAACACAAAATAAT  
AATACTATTATTTTTTAACTGGATTTTTGAAAAAGAAGATAAAAGTCTCA  
TTTTAGTAATTAAAACTCATTCCAGGTTAGTCCACTCAAACTTATATTC  
GAAAATTAAAACTTTGGGAGGCTGAGGCAGGCAGATCACCTGAGGTTGGG  
AGTTCGAGACCAGCCTGACCAACACGGAGAAACCCCGTCTCTACTAAAAA  
TACAAAATTAGCTGGGCGTTGTGCATGCCTGTAATCCCAGCTACTCGGGA  
GGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG  
CCGAGATCACACCATTGCACTCCAGCCTGGGCAACAAGAGTGAAACTCCA  
TCTCAAAAAAAAAAAAAAAAAAAAAAATTAAACCTCTGGAAGTTGAGTTG  
CAAATATTCATTATGCTCATTTTTAACTTGTATGTTTGGAAAATGTCATG  
ATGAAAATTGAGGTTGGGGGATGAGAAAAAAGAAAAACATCAACCCAC  
AGCCCATTCATTTTTAGCCCGACCCACAGCTCCGGGGAAGGACAGCAGG  
TCCATCCTTCACTCTTTCTTCACCTCTTTCCCTCCTTCTGGCTCTTCCA  
CCTCTAATTTGGAGCCCAAAAAAGGCACTGGGAAATGGAAAAGTCTTTT  
GTACGTGGTACTTGCCGGGGAAGCTGCCATGAAAACCTGGCCCCACGGTG  
GGGAGGGAATGCCCANCTGAGGCCTCGTGCCCATGCTAGGATAGACTCGT  
CCAAACATGTCAGGTGGTCTGACAGGGCAAGCANCANGAAATCATGTATG  
AGTATGAACTGATCTGTATGCAAGGGCGGGGAGAACACGCGGAGGAATGG  
GGCGTGAGAAACAGCACAGTACGTTTCTTTAGCAGCTGTCTCTGCTCAG  
CCATGGGAGGTACAGAGAAAGAGGCTTGGAGGCGTTATTTTCACTGTGA  
GATGTGAGTGTAAGAAAGTGCCCAAGACACAGTGAGTACCAGGGAGATGC  
CCTCTTTCCTACCCGAATGCAGAATGGCCACAGGCCTTAAACACACACA  
TGGGTCCTCAGAGGAGAGAGGCCTCCACAGTGACACCCGCATTCTCCCC  
TGGTCAGCAGCAGCAGGGCGAGTGCTGGGCCATCATGAAGCTTCACAGGC  
AATGAGCTCTCAGCAATAACAGGAACAGTGCCCTGGGGGACTGTAGCTGCA  
AGACCGATTTTTCTAAGATGGCCTCTGAGGACTCCGAGATACACCAGG  
CTGAGACTAGCTGGCAGCTCCAAGTTGTTGGTCAGAAGAGAACAGGAAT  
AGGGAAATTGGAATTACTGTTACTACAATTCCTTTACATCCGCACAACCA  
TGAGGTCCAGCGATTTTCTATTATTTTTTTTTTAAAGACAGGGTCTCAGT  
ATGTCGCCCAGCATAGAGTGCAATTGATGTGATCATGGTTCAGTACAGTAT  
TCACGTCCCAGGCTCAAGTGACCCTCCTGCCTCAGCCTCTCAAGTGGCTG  
GGACAGCAGTTGCATGCTACCAGGCCAGGCTTTTTTTTTTTTTTTTTTA  
GTTTCTGTAGAGCACATAGC

>Contig30

GGTTAACAATGGCACAGGGAAACAAACAGTTCCAGGTGCAGGGGCTCTAA  
ATCTATCATAAGATGTTAGGTATGGGGGCTCTGCCGGACACAACTCAAG  
GCTTTATGCTGTTATCTCTTGAGCGAAATCCTGGGAACTTCGTACATTGC  
TTGCTTCAGTACCTTATCAGTTAATCGGACTCTTTGATATGTTGGGAGTC  
AGCGTACACAAGTTAACTCCTTGAGGAAGGGGGTGGGTAAGGAGTCCTTG  
ATGTCTGGTAAATGAAGGAGCGAAATCGAGTTCCTCTGGCTTTCTCAGCT  
AAGGGAGAGCTTATTCATGTGGAACAAGGCTAAGTGATTAAGGGAGAAA  
GGGAGAGTCTGAAAACAAGGTTAGGTATTACAATGTCAATAAAATTGGTC  
TCCTTATACAGTCCTATGGTAGATTCTTTCCATCTTTAATCTCCCTCTA  
GCACCACCAGACTTTTTCTCTCTGTACCTTGAGATGTAAATTTTGCTATC  
TGAATTTTTCGTCTAAGAGTTGTTTCCTTTAATATGCAAATTTAGGGTTAT  
TTAGCTGACAACTGCCAAAGTAGTGAAACAAGTTATCAAGAAGCTTGAACG  
TCTAAGGTAGGAAAAAAAAAAGTCTTTATGAATCTATAAGATGTACTTCT  
ATTGGCATGCCTAATACGTCTATGTATTTACGTGTTGTGTACACAGTTTT  
TCACTACTGAAAATATATAGAGGAGTTCTAATTAATTGACTTAAGACAAT  
AAAAGCGCTTGAATCAAATACCTTATCAGGAAAAAGGAAAAGACAAGTCA  
AATGCTTGTTCAAGTTTATATAACTTAAGTAAATCTTTAATAAATAAGC  
TAGCTTTAACATTATTTGAAATGTCTTAAGAATTGCCAGCAGGTTCTGGG  
TTACAGAACTAGTGGGGGTGCAGTGGGGTGAGGGTTGGTGGGGTGGGGGG  
TGGTACGGGGGCTTTGTTTTTCTTGCTGCCCTTCTGGGTTGGGGAAG  
TGGGACAGCCTTGGCAGCACCCGAGCCGGCATGGCGTTAATAATGGAGG  
GATGCCAGACCAAGTGGCTAAGGCCGGCTGCAGAGCCAAGTTGGCATT

>Contig31

>Contig32

65/118

CTAAATTTTTAGTGAGAACTTGAGCTTGCTTCTGTGAGATATTTATTTT  
AAAACAGATTTGACACTTAAAATGTCTAATCAAGCCTTTTAAACCATGAT  
CTATCTCTTCAAATTTCTCAGATGCCACCATCAATAAAGAACTTTGTTT  
ACACAAGTAAGTGGTAGCAAATGGCAGGGTGTATCATTTTTTTTTTTT  
TTTTTTTTGAGACGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAGTGG  
CGCGATCTCAGCTCACTGCAAGTTCACCTGCTGGGTTACGCCCTTCTC  
CTECCTCAGCCTCCCGAGTAGCTGGGACTACAGGCACCTGCCACCACGCC  
CGGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTACCCTGTTAGCC  
AGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCCGCTCGGCCTCCCA  
AAGTGCTGGGATGACAGGCGTGAGCCACCGCGCCCCCGCTGTTTATCA  
TTTTTGCCTGATGAAATTTTCTTGCCACTACTCTGGATGGTTTGATAC  
ATTTAAATTGTGCTTCCAGGGTACAATTATCCTTTAAATCTATACCTCTT  
TCCTTTCTTTTATTGACAAATATAATGTTACACTTTTCTGTCAATTGCAGC  
CACACCACCGTAGACAGATCCCAACAGAGTTGTAATATTTTATTAGTTT  
CAGAGTTTCAATATTTTATCACTTTCAATACTTCATGTGCAGGAGTTTAA  
TTTGGTACTTCTTTACAAAATAAATGATGTGCTTCCAAGCATTCTTTTC  
AATAATTCCAATCAATGTTATTAAGTACTGAGTAATACTAGTATCTGTTTATT  
CATAAATTCACAGGAAATGCTTTTTTACTTATTAGTCTTTGGAATTCTGT  
TGTTTGTATAAACATCTTTCATGATGGCTTTGTGTCTACCAATAGCACTA  
TTGCCAAAAGGCACCTTTTTCTTGTTCTTTTACTTCACTGGTCCGAAGCC  
TGGTACCAACAACCTACCACACAGACTGGGAAATGAGCAATTTTGCCACGT  
GCCCTTAGCTATTAATGGTGGCACTCCATAACTAGCATCTTAAGCTCAAT  
TTCATGAAAGAAATGTGTTCTTATTTTGTACTTGCAGGCACTTTTTAAA  
CTTGTAATCTTTTATTTCATACTTTAAAATTAAACAGAGTAATAGAACC  
ATAGAAGGAAATCAATACCCACGAGTCCATACTGATATAAATAAATAGTT  
ACATAAATAAATGGGGGAGAAATAACAGCTCTTCTTACAGAAAAATTT  
CAATTAATAAATGAAGAAGGAATTAGGGAAATACAACGTTACCATTAAGC  
AACCACAGTAATAATCATTACAGGCAATATCCAAAAATAAATTCCAAAGC  
CAGTGGGCAAAAGTTTGAGGAGATACAGGATATTAACATAGTCTCCAAAT  
AGCTCATGCTATTTATAAATTACAAAAGGAAACATAACAACCTGTATAGTG  
AAGAACTCAGCAGACACCACCTTAGCCAAGTGATCAAGGTAAACGTCAC  
TAGTAATAGGGCTTGTTGACATACTGGACTCCAATCTGATACACTGATAA  
GGACACATGACTTCTGCAGTATTCTTACCAAAAACAGAATTCTAATGTAA  
TTAAGGAAAATGTGACAAAACCTATTCTGAGAAACATTCTATAAAACAA  
CTAACCAATACTTTCAAAATTGTCAAGGTCATAAAGACCAGGCGATGGTC  
ACAGATTTGAGGAGACTAAGGAGATACAACAATAAATACACAAATGGAA  
CCATGGCATTCTTGATTGGATCTTGAAACAGAAAAAGGATATTAGGAAGA  
AAAGCTGATGAAATTCTAATACATTCTGTAGTTTAATTAATAGTATTGTA  
CCAATATTAATTTCTTAGATTTGATCATTATACTATGGTTAAGTTTTTAA  
CATTAGAGGAATCTGGGAGAATGGTATATATGAACCTCCACTGTTCAATCA  
ACTTTTTAGTAACTATTATTTCAAAATAAAGTT

>Contig33

GGGAGCGGCGGCCACGCTGATCTCTAAAGCTTTAGACCACATTGGCTCG  
AGCATGGTCATGGCCGTTTCCTG

>Contig34

GACGTCTTAGCGCTATATTATAAAGAAATATTCACCTCCCTGCTGAGCTT  
ACAGGGTGTACCTAATGTCCAACAATATGAAATCTCTTCAATGAATTGCA  
GCACGTCCATATATAACCCACATGGAAGCTGTCTCTTTTCTCACCTTCG  
AATTCCCATGCCAAAGAGGGACCTCTTGGACTCAAATACATCTTAGCAA  
TATAGAAGATGCTGGAGACTTGTAGGAGAAGTGGAGAGGGTTTACAGTGT  
AGCCCCACAGAAAACAACCTATGACCCCATCAGTCACTTGTCCCTTTTTT  
CCATGCCTCAGTCTAGTCAGGAAACCACTAGATCCTGGATGGCTTCTTCT  
CCCTTCCCCTCTTTCTCTTCTCCTCTCCCTCCCTTGCTCCTCCTTCTC  
CATCACCCTCCTTACTTCCAACCAAACTTGACTAGCTCCAGTCTCAT  
CCCTCCTTATTGAAAATATTTTACTCAGCCCTCCTCCCCACTCCTGCC  
CAATCTTTATTCTTACCTACATCAGACTTCACCAAAACAAAGGCCAGGA  
TAATAAACAGGACAACTCTTCAAACACATTTTAATGACCATATTTTGT  
TATTTTGGTACAATTTGAGGAGTCCCAATCCCCAGGGAAGACTAACAAGA  
AGTTCTCCTAACAAAGGTGGGTCTCCCTTACTAAAACTCCTGTAATGG  
CTGAAAAGAGCATGAGGTTTTCTGCATATCATTACACATTCAATAGAACC

TCATGCAGCTGTTAAAAAAGATCTGTAGAGGCTATCTTGTGACAGAAAG  
GCATTGGAGATATACTGTTAGTGACAAAAATAGGTTATAAATGAATTTTT  
CCATGCATGCCTCTATATTTATAAATACACACACATAAAAGACAGGAAGG  
ACAGACATTAAACATTATAGTGCTTAAGATGATGCATAGTATAATAGTT  
AGGACCATGGCCTTTGGGACAGAAACTACAGCCTCTCTCCCACTTATCA  
GCCATGGGACCTTGGGCAATTTGCTCAGCCTCAAAGCCCCTGTTCCCTTA  
TCTGTGTGCTGGGGTGTGTGAAGAGTTAAGTGCAATACACAGAGAGAGA  
GAGAGTACCTAACATGTATTATGTGCTCAGTCAATATGCATCATAGTACT  
CATTTGTACATATGTTCCCTAAGTGCTTTATACGTTTTTTCCCTAAGTTGA  
CCATCTGTTTTTGGCATTATGAAACATAATGATCCTAACAAATTAAAATT  
AAAAACATAAAGAATATTTGCCCAAAAAAATAAAGAACATGAATTCCTC  
AAGTAGCCAAGGGGCCATAGACAGAAGTAAGCCCTTGGTGGGGCTTAGTT  
GAGAGAAGTCTCCAGAAGGTCTTTCGTGTGTTAAAGAAGAGGGTAACAGG  
GAGGAGGTGGGGAGAGATGTTAACTGAGTCTAAATGAGCACCTGGAAGAA  
GAGATGGGACAGGCCACTTCTGCCTGGACTCCCTGATTGTTAAGAAGAAT  
GAAAAAGAGCAGAAAGTCTTCCCTGAGCCCACTTCACTCCCTGACTTAAC  
CTAGTCTTTGCCCCCTTCCCTCTCACTCATGGCTACTTTCTGTGGTCACCT  
TGTTGTAGAAATGGATGTGCAGCCACCTCATCTTTTCTACCTCCTTCAC  
ATGTTTTAGATAATTTAATGTAGTAGAAGACGGTTACAGCAAAAAATTAC  
AAAAATCAAAATATCTCTGCTATCTACTGTTGCATTTCTAACCATCCCAA  
AACAGTAGCTGAAAACAGCACTCGTGGTCGAGCGCGGTGACTCATGCCTT  
TAATTCCAGATACTCCGGAGGCTGAGGCAAGAGAATCACTTGAACCCGGA  
AGGTGGAGGTTGCAGTGACTCAAGATCATGCCACTGCACTCCAGCCTGGG  
TGACACAGTGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAGCACTCGTG  
TATTTTGTTCAGATCTGTGGTTTGGGCAGGGCAGGGCTCAATGAGGACA  
TCTCGTCTCCGTTCCCGCAGTGTCAGGAAGTGTAAGTGAAGTGGAGGGT  
CACACAGAAGATGGCTCCCTCAAGTGGCCAGCAAATGGTGCTTACAATT  
GACAGGGAGCTGTTGACCAAGGGCCCCAATTCCTCTTCTATGGCCCCCT  
CTCGGGCTGCATGGGCTTCTTTACAGAATGGCAGCTGGATTCCAAGAGCA  
AGTATCACAACTACAGAAGAGTGGAGGAATATTGAAAGTTCACAGTCTC  
TTAAGACGTTGGCCAGAACTGGCAAAGCTTCATTTCTGCCATGTTCT  
ATTGATCAGTCACAGAACCTGCACCAATTCAAGAGGAGAACATATAGAGG  
ACATCTCTCAATGGGATAAGTGTCAACAAATTTGCATCTATCACAATCTG  
TCTTTTGGGTACAACTATTTCTATTCTCCATTATGCAAAATATACTCA  
CAACCTCCCAGGGGTGCAAAAGCCTCATCCATTTATGGCAAATGTGGCC  
CTTTTAATTTATATAAAATAATTTGCGGGGGCTTCCTTTATATTTTAAAC  
TCCCCTGC

>Contig35

GTGCAGAGAAGTGATTTAAAGCCCTTCAGAAAGAATGCTTTATTCCCCTG  
GAATTTGGTAACTTGCTTGGGTGTGGGGAGGTTTGTGAGCTTTCTCCACT  
CAAATTATCAGACCCTTTCCATTTAGTGGTAGACCATTTCCCTCGTCCAG  
GCCAAGGGCACATAGTACAGAGAAATAGGGAGTTGTTACCCAGGGAGAGA  
ACTTGGCTCTAAACCTGTAATAGAAAGGTCAGTTCTGGTCTGGAGGGTCA  
ATTTTGATCTTTGGCTCAGATCCAGGAATTGGAACCAAGGCTTTTGAACA  
TTTTAATGCAGGGGATTAATAAATGATACGAGTCATTACGAATATATT  
TGCTTAACATCTAAAGAGATCCCTCAAAACACTAGAAAAATAAGAACAA  
AAATCTAATAAAACAAAATTTGTTAAACACATTTACCAATTTTTTTTTT  
TGGTAAAAAATTCAAATGTCATAAATAAAGCTAAAGTTCCTCTTGATGACT  
CGCTCCTCTGCCCTATTCCACTCCAAGTAACCACTATTATCAGTCTTGCC  
AATACCCTTCAGACCTCTCTACCTCTATATACCATTAGAAGCACATGGT  
TTTGCAATTGAGGATGTGCAGTGTTTTGTTTTACGTAAATGTTATCACTCT  
GTTCTTGTTCCATAATTTGCCTTTTTCTCTCAATGATTTGCTTGGCTATC  
TTTCTATTTTCAGTAGCATCTCCTTTCTTTTTAACTTACCATTGTTTATTT  
AACCTTGCCTCTATCAACAGATATGTAGGTTGTTTCTAGTTGATTTTCAAT  
AAGTATTTATAAACAACGCATCAGTAGATGTCCATAAATTTCTTTACGGA  
AGATGGCAAGTAGTGGAATTGCTGAGCCAAAGAACATGTTTAAAAAACC  
AAAAAACTAGACGCTACCAATTTCTCTCCAAATGGCCATACCCACTT  
ACCCATACAGAGATGATTTGGAATCTGGCTTCTCACAAGGTGAGATGCC  
TTACAGTTTTCATTCTTCTGGCATGTCTTCCCTTTTGTATCTGAGAGAG  
CTGGCAGAATTGTGTCACTAAATCAAGGATAGAGGGTCAAATGACAGCTC

AAGCTCACAGGCACCTCTGCTTTCTTCCAGACCACCTGCTTTCTGCCA  
CCAGCTCTGTTCCATCTTATAGAATGGTTGCCACTTGGGTGCTGCTCCG  
ACAGCCATGTCATCCTTTGCACTGCAGTTATGAAGCAGACAGAGCTAGGA  
GAGGGGCTTTGCCAGCCTCTGCCCTAGCTTGGAGAACTTCAAAAAAGGAG  
GGTATTGAAGTTGAACTCCCCCAAAAAGGGGTGGTCCCCACACCTCAAAA  
AGTGGTGCCTCCGAAAGAAATGTAAAATTCGTGTGGGGGGGGGAAAAAGGT  
TATTTAGAAATTGTTGGCTTGTCTGTGCCGAAAGTATGTGTGGTTACGGGG  
AGTACGGAAATTCGAGGGGTGGGGGCGAGGCCGTGTGTCTTTAGCCCCG  
GGGTTTTCCCGTCGCATGTTTAAGGGGGGGGAAGAGGGGGGATGTTTTCT  
TTCCGCGAAGGTTTTTGAAGAACGGCGTGG

>Contig36

CCCCCACCGCCACTACTCAACCGGCCGTTACGAAACAACCTCGCCACAT  
CCACTAACCCGCTGGCTCACCACCCACCGCCCTCCCGATCCCCCAATCC  
AAACTCAACCCCCACCACCAAGCGCTCCCCCTCCCCACCCCTCCAGCT  
CAGCCCCAACCTACCACCAACCCCGACTCGCCACCGAAAACCAACAGCA  
AACCCAAATGCCACAAAACCAGTGTCAAACCCCTCCTTCCCATCAGTTT  
GGTGGGCCCCATCACCGCTTCCCCTGGGCCAGGCTCTCCTTTTGTGCGCTT  
GGAGCAGCAGACTGATCTCCAGCCTTCACTCACTTCATGTGGTAATCTG  
TTGTGTTTCATCACTGTGAGAATCTTCTGCATCCCCCTCACTACTCTGCTGA  
AAACACTCTAGTGGTTCCCTCATGTCTCATTAAATGAAAGTCTAGATATTAA  
ACGTAGAAGGCCCCAGCACAATTTGCCCTATGCCACCTACCTCTCTAATC  
TTTTCTCCTTACTCTGACAGACTCTCCGTCTGTCAATTTATGTATTCTTTT  
ATTGCTCTCTTCTACTTTTAGTATGAACTGGATTTATGGATTTTTTTAAC  
ATTGCTTCAAGTATGGAATAAAGAATTTTATTTATTTATTTATTTATTT  
ATTTGAGACTGGGTCTCACTCTGTTGCCAGGCCAGAATGCAATGGTGCA  
GTCATATCTCACTGTAACCTCGAATTCCTAGGCTCAAGCCATCCTCCTGC  
CTCAGCCTCCTAAGTAGCTATGACTACGGGTGTGCATCACCACATCTGGC  
TAATGGAATAAAATATTACAATGCCTAATCTTAATTTTCAAAATTTTAA  
TTACATTGTACCTAATGCCCATGCATTTACTTTTTTTCAGTGGGTCAATAG  
CCCTCACTTTTGGCAAAGGTCCCAGGCCCAAGGTAAGGCCTTACTTTTTCC  
AAACTCATCTTTTGAAGACATAAGTGCCTGTAAGTTGTACCACATTAGG  
TTCTAGGAATTTTTCATCAAAGACTTTATCAGACTATTTTCTCTAAGTT  
GAGAAAGAGCTGGGGGCAGAATATGGCACTGAATGACTGAAGAGAAGGCA  
CTGAAATCAGGCCAGAGGTTGCTGGAAAGAGCAATGAGGAACACCAGCAG  
CAATGAGGAGCCGGTGATGATTTTGGCTTCACAGGGAGGTGTGTACCACA  
CCGATTTTATCTCTACGTGGATGAACCACAGCTGTGCGCTCCCTTGCTC  
CAGGACATCACACTCTCCACATTCCCTCCCATCTTCCGGCTTCTGCTTCC  
CGGGGCCCTCATCTGCCCATCCTGGGTGAACACTGGTCCGTCAACTGCT  
GGCGTACCTTCCCCGCTCTGCACACCCTCCCTGGCCACCCACCCACTCT  
CACGGCTCGCACTGCAGAGGAGCCGCATCTCTAGCTCCAGCCCATCTGCC  
TCTTCTGAGCTCTAACTTCATGTAGGCGACTCCTGCCGGTGTGCTCAC  
AGGCCCATCATACTTCAAAGCATTTTCCCCTCAGAACACCATGTCCTGGC  
TGCTCCCTCCAGAAGATACATCTCTCAAGCACATCCCCGCGGTCTCACC  
TGGATGACTGCATTACCTTCTCCCATTTGCCCTCCTTTGGATGTA  
TATAGATTGTTTTAAATACAAATCTGATGTGCTTGCTCTCCTGCTTGAA  
ACACCTCAAACTGCCTTCAGGATAAACCACTGCCCTTGACATGTTTACA  
GGTTGCCCATGGCCTGGCCCTGCCCATCTCTTACGCTCATCTCATGCCC  
CTTGCCCTCGCTCTCTGGGCTTCTGCTCCCTAGCCCTCCTTTAGGTTT  
TCTAACACACCATAGTCTTCTAGTGTGGGGCTCTGCAAGTGCTGTTT  
CCATTGCCTGAGACATGAATCCCTCTCCCTATCTCTACCTGCACCTTCAT  
CTGATTAATCCCTACCCTTCTACTCATGATGTTGCTTCTCAGGGACTC  
TCTCTGACTTTTTAACTAATCAGGGTCTCCCCAGTATATATCTTCATAG  
CACTCTGTATTACTCCTTTCTTAATGACCACCTGCTGTAGACAGAATGTT  
TGTCTTCTCCAAATCATATGTAAACCTTCCACCAGAGCGATGATTAG  
AGAAGCCTCCC

>Contig37

GACTGACATTGAGAAGATATTAATAAGAGCACTAATGATGGGGATTGCAA  
CCATGTCTTTACTGACTTCCAGAAGCTTCTTACAGTAAACATGAAATCAC  
ATAATTTCTTCCACTTTCTACTGTTTCTTGTCTGGGCTCTGTCTGCT  
TACTGTCTAATATCTTGGCCCTTAAAGTTGCTAATCTTCCAAACCTCA

TTCCTGTGACTGGGCCCTGGTCTTG...CATGGGCCTTGAAGATAC1  
CTGTACACTTATCTGGAGCATCCAGTGCCTACCACCTGACCCAGATTCTT  
CATTGCGCTCCTCCCTCCTCCACCTAATGGGATTTGCTCATACCCGTGTG  
GGACCCCTCCCATTTTCCCAACTGAATACTTATCAAGACAACGCATTGC  
CATACTCCCTCGTACCCTGCTCTGGGCATCAGACTGAATGTTTGTTCCTA  
TTGAGGATCTGCAGCTGCATCAGTTTCCCCAGCACCGTCCAACCCCTTGA  
GCATGGCTAGTCCCTAAAGCAGAGAATTAGCCTTTCTATCCCTGCTGCTAT  
ACATGCTGGGACAAATAATAAGAAATGACAGCATTATGATAATGCAGG  
CTGCAGGAGGCAGGAGGCAGGAATCAAATTCGTGCTTATCAAATAGTGCT  
CCAATTCTTTGAATATTGGACTATAGAATATGTCATGGATCTATGCTCAG  
GTGGGTTCCCTATTACTCACTCCACTGAGGCCAGGTTGTGGGATTAGCTG  
TCCAAGAGGGAGTTTTCAGTCTCACAGCATAGGGTCATTCTGAGAATTACT  
GGCCACACTTGTGTGGAGACCTCCAGAGAACAGAATCTGGGTTGGTGCC  
ATGTACTTCCAGGAGGAGAGAAGTGGCAGGATGCCAGCCCCACAATCAG  
AGGGGAAGGGGCAGAGCCACATGTATGAAGATCCTCTCCCCAGTACGTGC  
CAATCACAGGGCTTCCTAGCTTTTGGGCCAAGGAAACAATGTGGGAAGCA  
AAAAAGGACAATTTCTCCTCCCTTTGCATGAAGACTGAGCAGTTTACC  
AGATTCCCAGGGAAACACCTTCCACTCTGGGTTGAATGTGAGTGAGAGA  
CATTACGCTGGAAACACTAGAAAACTATTTCTGAGCCACTCACCTTTAG  
CCCTAGAAAGTGTGGATTGTCTCTTCTGTCCTTTGCCACAGTAGAGACTGC  
TGATAGCATCAGAACTTGGGCTCTGGAATTAGACAGATATGGGTACAAAT  
CTGAGCTCTCTCACTTATTAGTGTGGGATGTAGAGCAACTTTTAAAATCC  
TTCCAAACCTCAGACTTCTCATGCATGATGTGAGGATTGTAATAGGGCCC  
ACCTAATAGGGGTTTTTGAAGAATTAAAAAAGTTATTCAATGAACAGCATT  
TAGCAAGATGCCTGACCATTGAGAAAATAACAAATTGTTTATTATTATG  
TTATTATTAAACATCTTCTCCTGCACCTTCTGACTGGGGGCATCGTATCAT  
CAGAAATACTTAGGATGGGATGGATTCTGTCATGGGCTGAGTCAAGGGTG  
CAATAATGGAGGAGTGAAGAAGGAAGAAATGGAGGCAGAAATCCCCAGGA  
GCCACAGATGGTACAAGGCTGAGCTAGTGCTGCAGAGCCTCCTTGAACA  
GCCACAGAGCTTGCATCTGGCCCTGGAGGAACCTCTTCTAGCTGGCAGGA  
CCAGCCACAACAGTGGCCAGGGGATTTCCAGGGCGTGGGCTCCTCAGGA  
GTTCAATTTGGACCAAGCCTGCCTGGAGAGGGGTTATAACAGGGATCCTTC  
CCTACTGGCAGGTGATTTACCCCTCGGTGAGAAGCTCAGGCATTTGTTTG  
ATGGAAGGTGGAAGGCCCTGTGCTGGGCCAGTGACTATCAGGGATGGGCG  
GGTGGCTGGAAAATAGCAAATAAGACAATATGATAACACAGTTAACCACC  
ACACTATGTGAAGCTACAATATGGGTATCTGTAATAGACAATTCCAATGT  
AGAGAAATAATTTAAGGTGTCACTTCTCCCGCCAATGCCATAAGCACACG  
GCCTCTGCCTGGGTTTCTCACTGTGGAATGTCCTCCTGGTCTCCTCATGC  
CCAGAGAGTGGGAAGTACTCCTACTTTAACACCGGCTTTCTGTCAATTC  
CNTGCAGCCCTCCTCAGCCCCCTCTGCACAGGGAGGTTTCTCCCTGCTG  
CTGCAGTGCTTTGTACTTGTAGTGGTACCTGCACACAGGTATTGGTGTG  
CTTGTCTCACCACCCTACATCACTGTAAGCTCCCCAGGAGCAGGCTTCCT  
GTTTGACTCACTGTGATCCTCCACCTCCCACCTGTAGTGCCTCAAGCA  
TTCTGTAGAGCATGGACGCC

>Contig38

GACTAATAAGTACTTCATTATTTGGGTATTTTCCAAGAACAACATATTGT  
AGGAAACCATTCTTTCTAAAAAAAAGTGTCTTTTAAAAAGGTGAATA  
ATTTTTGTCTAATTCAAAGTTTATTGAAAAGTTATGTATAAAACAAGGTA  
AAAGGAACAAGGAAATAAGGGAAATGTAAAGAAAATTATAGAAATAAAGT  
GGTATTTTTTGGTAAGAAAGCTTAAAGAGAAATAATTTTAGGTAAGAAAG  
AATCTTACCTAAAATTTTGTGCTAGAATAAAGTGAAGTGGCTAAGAAAGGG  
ATGTTCAAAGCTATTTATGACAAACCCACAGCCAATATCATACTGAATGG  
GCAAAAGCTGGAACATTCCCTTTGAGAACTGGCACAAGACAAGGATGTC  
CTCTCTCACCCTCCTATTCAACATAGTATCGGAAGTTCTGGCCAGGGCA  
ATCAAGCAAGAGAAAGAAATAAAGGGTATTCAAATAGGAAGAGAGGAAGT  
CAAATTTCTCCGTTTGCAGATGCATGATTGCATATTTAGAAAACCCCAT  
CATTTACGCCCAAACTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCT  
CAGGATACAAAATCAATGTGCAAAAATCACAGGCATTCTATACACCAAT  
AATAGACTAACAGAGAGCCAAATCATGAGTGAAGTCCCATTCACAATTGC  
TACAAAGAGAATAAAATACCTGGGAATACAACCTACAATGGACATGAAAG

ACCTTTTCAGGGTGAAC...GCAAACCAC...CTCAAGGAAATAAGAGAGGAA  
ACAAACAAATGGAAAAACATTCCATGCTTATGGATAGGAAGAATCAATAT  
CGTGAATGGCCATACTGCCCAAGTAATTTATAGATTCAATGCTATCCC  
CATCAAGCTACCATTGACTTTCTTCACAGAATTAGAAAAAACTAATAGCC  
AAGACAATCCTAAGCAAAAAGAACAAAGCTGGAGGCATTGCTGCTACCTGA  
CTTCAAACCTATACTACAAGGCTGCAGTAACCAAAACAGCATGGTACTGGT  
ACCAAAACAGATATATAGACCAAAAGAACAGAACAGAGGCCCTCAGATATA  
ACACCACACATCTACAACCATCTGATCTTTGACAAACCTAACAAAAATAA  
GCAATGGGGAAAAATAATTCCCTATTTAATAAATGATGTTGGGAAAACTGG  
TTAGCCATATGCTGAAAACTGAACTGGACCCCTTCCTTACAACCTTATAC  
AAAAATCAACTCAAGATGGATTAAAGATTTAAACATGGCTGGGCATGGTG  
GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGATGGGTGGATCAT  
GAGGTGAGGAGATGGAGACCATCCTGACTAACACAGTGAAACCCCTGTCTC  
TACTAAAAAATACAAAAAATTAGCTGGGCATGGTGGTGGGCGCCTGTAAT  
CCCAGCTACTTGGGAAGCTAAGGCAGGAGAATGGTGTGAACCCAGGAAGT  
GGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCTAGCCTGGGCAAC  
AGAGTGAGACTCCATCTCAATAAATAAATAAATATGGAACTCTCCCAACA  
CAATAATAAGACAAACCCCCAAATGTTTTAAATGGGCAAAAATATTTGAA  
CAGACACTTCACAAAAGAGGATATGTAAATGGTCAAAAAGCACATGAAAA  
GATGTTCAACACCATTGGTCAATCAGGGCAAAGAAAACTAGAACCACAATG  
AGATGCCTCTGTACACCACCTTAAATGTCCAAATTAAAGAAAAACAAGTTTT  
GGCAAAGTTGTGGAGCAACTGAAATGCTCGTGTATTGCTGGTAGAAAAAC  
AAAATGGCATAACCATCGCAGATAATTTGTTGTGAGTTTCTTACAAAGTT  
AAACATATACTTATTGATATGACAGTTCCATTCCAAGAGAAATGAAAAA  
TAAGTCCACACAAAGACTTGTACCTGGGTGTTTATGGTAGCTCTATTCAT  
AATTGCCAAAATCTGGAAACAAATCAAATGTCCATCAGCAATGGAATGGA  
TATACAAATTTGTTGATACATGTACAATAGAAAACTACTCTGCAATGGAG  
AGAAATTAACCATTGACAAACACAAAAACATGGACAAACCTCAAAAACAT  
TATGCTGAGCAAAAGAAGCCAGACACAAAAGACTGCTCAGCGCATGATTC  
CATTCAATATGAAATCACAGAAAGGGTCAGTTGAAGGTGCAGAGACAAAAA  
GTAGATCTGCAGTTGCCTGGGGATGGGGTGGGAGGTTGACTGCTCTGACG  
CGTAAGGAAATTTGGGGGTAGGTGGGGGATGGTGGGAATATTTTTTGAAT  
TGAATTGGGTAATAGTTTTAATAGGTAATAATATTGGACCCACAGTATTT  
GAGATAGGTTTCAGTCAATTTAGACAGTTTATTTTGCCAAGGTTAAGGAT  
GCATCCGTGACCCAGCCTCAGGAGGTCCTGACAACCTGTGCTGAAGGCAG  
TCAACATACAGCTTGCTTTTATTCATCTTAGGGAGACATAATACATCAAT  
CAATGCATGTAAGGTTTACATTGGTTCAATCTGGAAAGGTGAGGGAACTT  
GAAGCAGGGAGCTTCCAGGTTACAAGGTAGATTATTCTCAACAGAAAGGA  
ATGTCTGGGTTATGATAAGCGGTTGTGGAGACCAAGGTTTTATCTTGTAG  
ATGAAGCCTCCGGGTAGCAAGCTTCAGAGGGAATAGATTGTCAAAGTTTC  
CTATCAGACATAAGGTCTGTGTTGATGTTAATGCTGGTCAGCTTTTCCTG  
AATTCCAAAAGGGAGAAGGGTATACTGGGGCATGTCCAACCTTCCCTTCC  
ATCATGACCTGAACTAGTTTTTTCAGGTTAACTTTGGAATGCTCTTGGCC  
AAGAAGAGGGGTCCATTGAGATGGTTGGGGGGGCTTAGAATTTTATTTTT  
GGTTTACAGTGAAGACTTTTCAAGCTAGACACTTAAATGAGTATGTTGCA  
AAATGGCAATTTCTTAGCACGGC

>Contig39

GACGTCCTAAAGAAATGCTAAGGTAACCTCAATTAACCTATGCTAGAAAAGA  
GAGTTAAGTATTTAGGAGGATTTAATATGGTGTTAAAGTTGTGAAAATCA  
AAATGGAGACACTAATGTAAAGAAAACCTGATAAATGGAGCCAGGGAAG  
GCCATGAAGAAAGAGTTCTCACACTTGTATCCCTGATCATGAAAAAGACT  
CTGCAAAAAACAAACCTTGACAAAAGGCCATTGCAACCTTACACAAAAA  
ATACTACTTTAAAAGGACATGTGCCAGCAACTGCCTGTCCAACCTCAGA  
CTGGCAATATCTTTGTTATTGATCTTAGTAGCCCAGCATAACTATTTCAA  
AACAGTGATGTAATGCTCATTTTTTTTTCTTTTGAAAACCTTTGTCTTCCT  
GTAAAAACCTTTGTCTTCTTTACTTACCCTGAATATGCACAGAGTTTACT  
ATGGAGTGCATATTCTGTGCAATGCTCTATTTCCCAAACAAACATCATT  
TTCTTTTAGAGAGCCTCTCTCTGTTTGTGATTTAGGTTGGTGATGTAAAG  
CAATGGCATAACTGAACACTGATTCAAAGAAAAGTGGCTTTTCTCTTTGT  
TGTATTTAAAAGAGGCCCTTATAAATAGGATAGTAAGATTTGTAAGTTGAA

CTTAAAGCATGAAGAAATTTAGGGGCCAGGCAGGGTGGCTCACACCTGT  
AATCCCAGCACTTTGGGAGGCCAAGACAGGAGGATTGCTTGAGCCCAGGA  
GTTCAAGACCAGTCTGGTCAACACAGACCTCATCTTTACTAAAAATAAAA  
AAATTAGGCCAGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA  
GGCCAAGGCGGAGGATCACTTGAGGTGAGGAGTTCGTGACCAGCCTGGT  
CAACACGATGAAACCCCATCTCTACTAAAAATACAAAAAAATTAGCTGGG  
TGTGGTGGCGGCACCTGCAATCCCAGCTACTCGGGAGGCTTCAGGCAGG  
GGAATCACTTGAACCTGGGAGGCGGACATTGCAGTGAGCTGAGATAGTCC  
CACTGCACCTCCAGCCTGGGCGACTCAGCAAGACTCTGCCTCAAAAAAAA  
AAAAAAATTAGTCAGGTGTGGTAGCACACAGCTGTGGTCCCAGCTACTC  
GGGAGGCTGAGGTGGGAGGATCATCTGAGCCAGGAGGTCAAGGCTGCGG  
TAAGAGCTGAGATTGTACTACTGCATTCCAGCAGGGGCTACAAAGTGAGA  
CCCTGTCTCAAAAAAGAAAAAGAAAAAGAAATTATGTTTTTAAATTTA  
TAATTATAATAAATTTAATTACATAAATTTAAGCTCAAGTAATTGTAAAT  
ATTCTTTCTGTGCACATAAGTTATTCTTGTATTGACCCACAGGAGCTGG  
CCATTCTTCAAGTCAGAAGGCCTGAGAGAGGAGCTGCCAGGTGGTCTTC  
ATGGGGCTGTGCGGCCAGTCATCCCCACAGGTTGACAACTCCTTGTGTAC  
TTCATCCTCGTTGGATCCTCTGTATCCCTGACGATGAGCAACTGTGAGGC  
CCGTTTCAGCAGCTGAGTTCAGTCAGGAAAACATCCACCCACCCACCACA  
CGCTCACACTTACACACACATTACACATGCACACAGTTCTGGCTCCGA  
AAAAAGAAAAAAAGCAATTTAAATAATTCTGATCCTTTGCTTATTT  
CCACAACTCCATGAAAATTGTACATTGTCCAAGCAACATTTCTTAATAT  
TCTCTTTTTCTCTCATATCCATTTTCTTACTGCTGTCTCCACCTTTCTC  
TTCCAACTCCCTGTTAAATCCCTGCCCCAGCGAACTTTTATTCAATTT  
TGTGGAATGGAGGCTGCTCTGATTTAAATTAAAAAATAAAAAATCCC  
TACTCCATGTCCCAGATCCCTAGTTGTTTTTTGTTTTTTGTTTTCTGAG  
ACAGGGTCTTGTGTCTTCCATGCTGGAGTGCAGTGGCATGATCATGGCTC  
ACTGCAGCCTCAACCTCCTGGGCTCAAGTAAATCTCTTGCGTCAGCCCTC  
CCCAGTAGCTGGGAGTTCAGGTATGTGCTACCATGCCTAGCTAATTTTTT  
CTTTTTATTTTGTAGAGACACGGTCTTGCCAGGTTGCCAGGCTGGTATA  
GAACCCCTGGGCTTAAGTGATCCTCCTGCCTCGGCTTCCCAAAGTGCTGG  
GATTACAAGTGTGAGGCACTGCACCCAGGCTGGATCCCTGCATTTTACA  
GATTTAGCATCACAAAAGTCTAAACAATTAGACTGACTAAGGCAGAAGT  
CCCTTATGACAGCAGACATAAGAAGGAAAAGGCCAAAACACTGTGTTAAA  
AATTATCCAAATGTGAGGAAAAGGCCAAAGAGAGTAGGTGTGCCTTTTTAG  
TGTCTAAGCTGCCTGCCCAAGGGGCATCTGATGCTCTCAGGCAGGAGTCC  
ACAAATTTTTTTTTGTAAAGATCAGATAGTAAATCTTTTCAGCGTGAAG  
AGCATGAGGTCTCTGTCAAAATACTCAACCACCATTAACATGAAAGC  
AGCCAACAGACAACACATGACAAATGAGTGTGGCTGTGTTCCAGTAAATC  
TTGATTACAAAACAGGCAAGAGGCCAGAGCTGACCCATGGGCCATAGTT  
TGCTGACCCCTTCTGTAAAGGAAAGTATTTTTGTTTGAAGTGTGTTTAC  
CATTGATTGAACACAAGGCTCTGTAGAGTTACTTGTTAACTTGACAGA  
TTGATGAGTGGCAAGTAATTTTTATTACCAGAATATANNATTATTCTGT  
TCAGTAGATAAGATAAACCCACTGTTATATTACTGTCTTGTGTTAGAAATG  
GACTTTGATTCAATTTTTTACAAATTCATATTATTGCCCTAATTTGTATA  
TAAGTATGCTTCTTTTAAAAATATATATTTTTTAAATAAATTTGAGACAGG  
GTCTCACTAGGTTGCCAGCCTTTTGCTATAATGAGAGCATAAAGTGAAT  
TTCACACTTTAGCCTAGTGCATAGATGGGATTACAGGCACAAACCACTGC  
ATGCAGCTAACTTTGCTTCTCATTCAGCACGTTCTATTCCNNNGNTTTT  
CATATACGCGTCTCTTAATGC

>Contig40

CGCATTGAGCCCAAGTTTTCTTCAGTGTTAAGGTTTTTGTACTCTGTGC  
CCAAATGTCCTTCCAAAAGGTTAAGTTTTTTTACCTTCCTGCCAACATT  
ATATGAAAGTGTCCACTTTTGTAGACTTTTACCAATGCTGACTACTTTTG  
GTTTCAAAAAGCTCTCAGTAATTTTCTATTAATTACTTTTACCCTTTTT  
TATTGAGGGTGTCAACTTTTTATTGTTAGCATATTCTCTCTGGGCTCCA  
TTGGACGCCTTGGCAGCTTTTGGTAGTAGGTGCCTTTAGAAAAGTCCTT  
CTCGTCTGGCCCTTTCTGAGCAAATCTAGTGAACAGAATTGGCTCCATGC  
TCAGCAATTGCTTAATACGGTTGATCCAGGGCCTAGGACTCATTCTTCAT  
TACCATCCACTTGCATTGTCTTAAAGCAAGGCTCTATTAATTTAATTGG

CATTTCCTGTCCCAGCTCTCTTAGTTTTCATTAAACAAAGGCTTTAGAAAAAC  
TCCCAGTAGATGCCTATGTTGCTTCCTTTTAAAAAATTTTGGAGCTGTTT  
CCCTAGCCTAACCTTTTCTTCAGGGCAGGAGTTAAGTCCCTTCTACTGCA  
TTCTGTGAAGATGGTGAATCAAGAGGCAGGGCACCTGTTGCTTTGTGAA  
ACAGTCCACTCTGCAGCTGGGCAGCTCTGTTACTAGAATGTTCTCCCTTC  
TGGGGAGCCAATATTTTGATGTCCTCTGTGAATCTCATCTGCTTATCCCA  
TCFCTTTATGTCCTTGAAGATGCACAGGTCTGACACCACGAGGTAGCCCT  
TAGAAATTTGATGGCATTCTGATGTGTCCCCAACTCTTCTCCAACCACT  
CCTCCCAGAGCTTGTTCCTTAAGCCCCCTGTGGAGCTGATTGCTTTCCTC  
AAGGCAGCTCAGTTTTTCCAGTTTGCTCCTGGTGGTCTGAAATATGAT  
TGACTCCTGAATACTCCAGGTGTGAAGGAGAGTGGGGGTGGCCTTTCTAC  
TTGTCATGGCCTGGGTTTTAAGTTGCTGTCCAGTGGAGCAGAGGTGACTT  
TCCCAGTGAATACTATTTTTTCCCTCTAAATCCTTAGCAATTTTGTCTC  
CAGAGGCAAGACCTGGCCAAACCATTGTGTTGAGGATTGAATCAAGAAT  
GATTGAGGAGATGACAGTAGTCCCCCTCATCTGAGGAGGGCGTGTTCCT  
AGCCCCCTCAGTGAATGCCTGAAACTGTGGATAGTACCCAACTTATATGT  
CTATGATTTTCTATAAATTAATACATGCCTGTGACAATGTTTAATTTAT  
AAATTAGGCAAAGAGGGCCAGGCGCAGTGGCTCAAGCCTGTAATCCCAGCA  
CTTTAGGAGGCTGAGGCTCACCTGAGGTGAGGAGTTCGAGACCAGCCTG  
ACCAACATGGAGAAACCCCGCTCTACTAAAAATACAAATTAGCTGGGC  
ATGGTGGCAGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG  
AATCACTTGAACCCGGGAGGCGGGATTGCGGTGAGCTGAGATCGTCTCA  
TTGCACTACAGCCTGGGCAACAAGAGTGAACTCCGACTCAAAAAAAAAA  
AAAAAATTAGGCAAAGAAAGAAATTAACAACAATAAGTAATGAAATAGA  
ACAATTCTAACAATATACTATAATAAAGTTGTATGAATGTGGTCTCTTT  
CTCAAAATTACCTTTTTTTTTTTGAGACAGGGTCTCACTTTATTGCCCAGG  
CTGGAGTGCAGTGGCAGCATCACAGCTTACTGCTGCCTCGACCTCCTGGG  
ACCAAGTGATCCTCCCACTTTAGCCTCCTGAGTAGCTGGGACCACAGGCA  
TGCACCACTGTATCTGGATAATTTTGTATTATTTTGTTCAGAGAGAGG  
AGGTCTACTATGTTTTCCAGGCTGGTTTTGAATGCCTGGGCCCAAGGGA  
TCCTCCTGCCTTGGCCTCCCAAAGTATTGGGATTACAAGCGTGAGCCACC  
ATGCCTGCCCCAAAATTATCTTATTGTTCTATACCCACTCTTCTTCTGT  
GATGATGTGAGGTGATCCATTGCCTCCTTGATGAGATGAAGTGAGGTGAC  
TGATGTGGGCATAGTGATGCACTGTTTAGGCTGATATTGGCCTGATGATA  
TGTCAGAAGGAGGGTCATCTGCTTCGGTGATCCTGGATCATAGAGTCATG  
ATGATGTCAATGGTTGGATGTGAGGAGCAGACGATGTCAATGACTAACGA  
TAAGCTGGACAGGTGGGATGGTGGCACAAGATTTTATCACGCTACTCAGA  
ATGGAGCACAATTTAAACTTCTGAATTGTTTATTTTTTGGAAATTTTTCAT  
TAATATTTTTGGATTGCAATTGACTGTGGGTAACTGAAACTGTGGAATGT  
GAGACTGTGGAAGAGTGAGGGAGTACTGTATTATGGAAGTGAATCTAT  
TCGGTAGGGGAACAGAATTCACATTTGTGGGGCCAGGTCTCTGCATCTG  
TAGGGATCCAATTGTTTCATTTCTCGTTGTAGCAAAACTTGGCTTTGGA  
ATCAGACAGATTGATGTTTGCTATCATTCTAAATGGGTGCAGCTACACTT  
TCCTCAAGAGGTAGTTCTGAAAATTTAACAATGTGAATTTCTTGGTAA  
AAAAAAAAAACCTCAAAATATTAGTTTCCTTTCTTTGTGTCTGATGT  
ACTCCATCAATACTGGGAAATATGTGTCTCTCATAGAAATGTCATGGAT  
CTTTGTAATTCTGATTATCCACAAACCTTGGGGATTAGCTGTTTCAATGT  
TCCTATTTTACAGATAAGAAATGGAGCCTGTGGTAAGTTAAGTGAGTTA  
CTCATGGCTACTTAACATAATTTTACTAGGTGATAGGCCAGAGCTAGAG  
CCCAGGTCACTTCTTATCAATGCTCTGCCTTGTCTCTGTGCCTTCTGT  
CTGTCTGTATGTGTATGTGCCTGTTGACAGTAAGGCATAGTTTAACCCAG  
TAGAACTACCGGTTTGTAAATGAATCCACTTGTAATGACTGACCATTCA  
AGGAACAAGTGTTTTTCTATGCTTGACACCTGTTTTGGATGCCAAAAG  
GATACAAATGTAACCTTCAGACACTCTGGGCCTCATTTTGCACCTATTAGC  
ATGTCCAAAATTTAAAAGACTGACCACACCAATATTGGTGAGGATGTGG  
AAGAACGGGAACCTTTCATACACTGCTGGTGGGGATGTAAAATGGTACAAT  
CCCTTTGGGTAAACAGTTTGACAGTTTCTTAAAAAGTTAGACATATATATT  
TACCATACTGACTCAGCCCTTCCACTTCTAGGTCTTTACCCCAAGAGAAATG  
AAATGCTGTCTTTTTACAAATGTCTATACAGGAATGTACATAGCAACCTT  
ATTTGTCATTGCAAAAAACAGAGACAATTCAACGTTGTCAAGAGTGAATG

FIG. 4 (18 of 61)

72/118

GATGAGCAAGCTGTGGTCTCTATGCA...GGTATCCTACTCAGCCAG  
AAAGATATGGCTAAT  
>Contig41  
GACAACAATGTCATGCATAAGATGACGATGGCCTGGGTGATTGATGCAAA  
CAAGGATAAAGAAAATAATCAATTTTGTCCCCATTTTCAAAGACAGATAG  
CAGCAGCAAGAGTGTAAGTCTGAGGAAAGTCATATTCCTTCCTCCTACAA  
CATAGCACACACACTTACAAAAACAATACACAGACTCCTGGCCAATGGAC  
TTCAAAACTGAGGAGGATCATTAAATTTAAATGTTACCGCTGCATGAAA  
TCTCCCTGGGTCTGCCCCTCCCTTCCCCACCCTCCTCCACTTGGGCCGGG  
GCACAGCAGTGATTCTCTCACCTCTCAGAGTGAGCCAGTGTTGGCTGCAT  
TGAAGGCTCCAGATATGCAAACAGGGCAGATATTCCTGGACCAGGGTGCA  
CAGAGTGAGGCTCCAACGCACCCTATTAAGTGCATGAAGGATGAATGAGC  
CTCTGGTATGGGCTGGGACAGAAAAAGGATTCAAGGGGCCAAAAGGGT  
TTGGGTGGAACCTACCAGGAGCGGCAGTACAGACTCCTTGGGAAGGTGGC  
CATGATTTAGCCACATTACCAATAGGATAATCTGGAGAATTCCTAGCT  
TGAGTTTCTGGGAGAAAGCAGATTTCTGGATTATCTGGTGACAGGTAACA  
GGGCCGAGTTTCATCCACAGCCACCTGCAGTGTTAGCACCTTAAGCTGAGT  
TCCTTGACCAGGATGCTGTACGCCCAGTCAGTGTGAGACGGTTCTTGG  
CTGAAGGACTGAAAAGCTTGGGTAAGTGACTTCACCTAAGCCTCTATCTC  
TTGCTCCCGTAAGTCAGGGCTCATTGTGGCTCCTTGCAGGCTTGACTTCA  
GGGTTAACAGAGAAAATGAAGGTACAAGTGCCTTGTGAAGCTCTGAAGCTC  
CAAACAGTCATTCTCAAAGTGCCGTCCACCAGTCTAGCACATCAGCATC  
ACTGGAAGCTTGTTTGAATGTAAATTATCAGGTCTCCAGAGCTATGTA  
TGAATTAGAACTCTGGGAATGGGGCCCTGCAATCTATTTCAACAGGTCC  
TCCAGGTGATTCTGATGCAAGTTAAAGCCTGAGAACTCTGTCTATACA  
AATGGATGTCAACTCAAGCTGCTCTTCAGAATCACCTATAGCACTTGTTT  
ACCCGAATCCCTGAGAATGGAGCTTCAGGACTGCTATTTCTCAAAGTTTG  
CCTGGTGATCCTGAGATGGGGTTTGGGGGACAGAGATCCAAGGTGCTACC  
AGGTGTGAGGAATTGTTAGAAGGCAAACCTGGCTGTCATCTAGGGTGCTT  
AAAGGGTACAGATCCTAGGATTCTGCCTCTTACAGCTGAATCAGACTTTC  
CTAGAATGGGATTGCTGTCCAATGGCATGCCTCCTGGGTGACTCTGATGT  
ATAGCCTGGGCTGGGAACCACAGAGGATTATCTTCCATTGACCAAGCTG  
ACAACTCGCTTAAGGCTCTGAGTTTCACACTTGATTTTCTAGCCCCTGT  
CCTTCCATGGATCACCTGCCCCCTTCCCTCCTAATCAGGAGCACAGTCAG  
TGGATGCATAATGTGGCCTCTCCTTGGCTGCAGGGAACAGGTGGAAATG  
TGGCCATAGGTGTGACAGGGCTGCCTGCCATGTATTAATAGCTACAGATT  
GAAAGATCCAAGGACAAGAGACTAGAAAAAATTTAAACAGCCAAGCAT  
TGGCCCAGTAATGGCATTTCAGAAATCCACCAAAATATTAAGATGCTTTT  
TGAAAAATATCCAGAGCACTCATGTAAAAGTGCTTAATTATTAATAAAG  
CTGACATGTGTTGGGTACTTCTGTGGGTCTGGCACTAGGCTAATTATGT  
TTTTAGGAGTTGACTCAAATGCTCCCTGTCATAATTATGTGAAAAATAT  
AATTATTAGCTCCATGGTACAAATTAAGGAGAGGTTACATAAATAAAG  
GAATGATACTCAAATTAGTAACCAGAGCCCATGCTCTTAAACACTATGCT  
ATTATTTGTGGACTCTTACATAGGTGGCAAAAGTCAAAGGCTAGATTGAC  
TTCTGTCCACTTCCAGCCAAGATGAAGTACAAGATTACAGATACACCCTTC  
CGCATTAAACAACCTTAGGAATCAGACAAAATATACAAAGCATTGTTTGTT  
ACACATTGGATAACAGACAGCACTAGATAGTCGTGTCTGAGAAAAGCGGT  
GAAATGAGCTGAGTCTTAGAATTGCCCCAGTTTACTAAGGGGCATAGTAA  
GGGCATAGCTGCAGCACAAAGAAGCAGAACCCAACAGAGACTGGCGTTCA  
CCTGAGTTGAGAAAACCAAGTTGAAAATTTAGGAACACTAACACAGATAT  
GTAGGCAAGAGTATCAGAGAGGAGACAGTTGTAGGGAAAAAGAGAGCTTT  
ACAGAGAGACAGCGAGAGCTCCAGAGACCCGAGAAGATTGCCCTGACGT  
CACTAGCTGAGTACCGATCAGTGCATACATGTAAGGATATTACTCAATAT  
GTGGAAAAGAACAGAAGGAATGATGTCAAAGCTCACCCAAAGACAGGAA  
TCATTTATGTTTCCACCAGCCAGAGTGGAACAACCTTGTAACGCATATGG  
AGTACTCAAACGAATATTTCTCAATAATAAGTTCAAATTAAGTGAAGT  
AAAGCCTGCCCGCTTTGTCTGGACATGCCTAACAAAGCTTTGAGGGAAGC  
CTCAAAAGAATGAAACCGTGTCCAAGTAATTTAACTGTGTCCAGAAAAA  
AATTCAGAACATTTAAATAAATATTAATAATATGATCAAACCCAGCAAGG  
TTAAATTCAAATGTCTGGCATCCATTAAAAAATTACCAGCCTTGAAAAAT

FIG. 4 (19 of 61)

73/118

TGGCGGGAAAATATTA: .ATAATGAA. .3AAAAAGCAATCAACAGAG/  
AGGCCTAGAAAGTATACATATGATAAAATTAGCAGACATTAAATGGTTAT  
GATTAATTTATTTTATATGTTAAAGAAGGTAGAGAAGAGCATAAGCACAT  
TAAAGAGAGACAGGAAAGTCCCAGTACTCACACAGGGCCAGGAGCAGTTT  
TCACCAAGTCAGGTGGGAAAACCTTCATATTTTCATGGAGCATTGGTAGAGTA  
CACAGTGTCTTGCTTAGTAGAGGGATAAAATGCTGTTCTGTTCCCGCCTA  
ACCCATCTTGAAAGAAAATCTGAAAGGATCAAACTGTATTCAAGTAACCT  
AATCACATCCCAGCACACAGCTCGACTAGTTATAAAAACACAAAATATTA  
ATATCTAGAAACACAAAATAATATCTAGCACCCAACAAGGTAAAATTCA  
CAATGTCTAGCATTCAATTGAAATTTTCTAGGCCATCAAAGAAGCAGTAA  
AATATGACCTATAAGGCCGGGCACATTGGCTCATGCCTGTAATCCCAGCA  
CTCTGGGAGGCCAAGGTGGGTGGCTCACCCGGAGGTCAGGAGTTCAAGAC  
CAGCCTGGTCAACATGGTGAGACCTCATCTCTACTAAAAATATAAAAATT  
AGCCAGCATGGTGGTGGCGCCTGTAATCCCAGCTACTCAGGAGGTTGA  
GGCAGGAGAATCGCTTGAACCTGGGAGAAGGAGACCGCAGTGAGCCAAGA  
TGGCACCAATGCACTGCAGCCTCATTAGAGAACATCGGGAAG  
>Contig42  
GAAACTAAAGGCTTATTTAAAGCGCGAGACCGTGGCGCCTTTGGACTGGA  
CCCTTTCTAATGATCATTTAGTATCAGGCTATGTGGGAGTTGACCGTTTT  
GCATAGCCTGAAAGCCAACAGTATCACTCCTCCTTAGGTGTGGCAGAGA  
TGTGAGAGAAGGAGACTGACAGTCTGTGGGTGTGTATGEAGTGTTGGGG  
AAGCGAGGCACAGGGGACAATACTGTGGTGTAGAAAACCTAGTCTAAGGTA  
GCATCAGGAAATTCATGAAACCAAAATGAATTTTATAACAGCACAAAGACA  
TTATTTGTTTTTGCCTCCCTCTCATTTTTTTTTTTTTTTTGAACAGAGTC  
TTGCTCTGTTCATCCATGCTCGTGTGCAGTGGTGCAATCTCGGCTCACTGC  
AACCTCCACCTCCAGGGTTCAAGCAATTCTCATGCCTCAGCCTCCTGAGT  
AGCTGATTACAGGTCTGCACCACCCCGCGGCTAGTTTTTGTATTTTTAG  
TAGAGATGGGTTTTGTAAATGTTGGCCAGGCTGCCCTGTCATTTTTTTTT  
TACTAGTGTCCAGTGGAGTTTTTATAGGGGCTACATAACATGATACTGTCA  
TTAATCTAATGGCTAATGAAAGGGATATGTATATGTTTTTGTGTTTAAAA  
CAAACCTTCTTGGGGTCCCTCAATAATTTTTAAGAGTATAAAGGGTCCCTG  
AGATCAAAGAGTTTGAGTTCTGCTGGACTGGGACAGTGGTTGTCAACCCA  
GATTGTACATTAGGTCATCTGGGAAGCTTTAAATAGTACTGATGCCCCA  
ACCTTACCGCAAACCAATTAAGCCAGAATCTCTGTGGATGAGAAGTCTTC  
ATTGTCATCATCACCATGACCATCATCATTGTCACCGTCACTACACCATT  
ATCATCATCATCATATCATCTTCATTATCATTGTTAGTATCTCCATCACC  
ATCATCAGCATCACCATTATTATCATCATCATCATCCCCACCATCATCCT  
CATCGGAACCTCACCTGCATGGAGGACAATCCACTATGCATTAGGTGCTA  
TGCTATTTGCTATACTCCTTATTCTCACAACCTGCCAGAGAGGCTGATAT  
TATCTCACTTTATAACAGGAGGAATCTGGATCGGAAAAGTTAAGGTAAGC  
TAATTCACAGAGCGAGAAGAGATAGAGCCAGGATTCGAAACCAGTTCTCT  
GCTCATCAATGTTCCAGTCTTGCCTTGCCTATTGAGAACCTCTTTAGTTAT  
GCTTTTCAACCCCTCCAACACCACAGTAAATTTTTCTTTTTTTAAAAAAT  
TATACTTTAAGTTATAGGGTATATGTGCATAATGTGCAGGTTTGTTACAT  
ATGTATACATGTGCCATGTTGGTGTGCTGCACTATTAACTCGTCATTTA  
CATTAGGTATATCTTCTAATGCTATCCCTGCGGCTCTCCCCACCCCATG  
ACAGGCCCTGGTGTGTGATGTTCCCCACCCTGTGTCCAAGTGTTCTCATT  
GTTCAAGTCCCACCTATGAGTGAGAACATGTGGTGTTTGGTTTTCTGTCC  
TTGTGATAGTTTGCTCAGAAATGATGGTTTCCAGCTTCATCCACGTCCCTA  
CAAAGGATATGAACTCATCCTTTTTTATGGCTGCATAGTATTCCATGGTG  
TATGTGTGCCACATTTTCTTAATCCAGTCTATCATTGCTGGACATTTGGG  
TTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCCACAGTGAACATTCATG  
TGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCC  
AGTAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAG  
GAATTGCCACACTGTCTACCACAATGGTTGAATTAGTTTATAGCCCCACC  
AACAGTGTAAGAGCATTCCTATTTCTCCACATCCTCTCCAGCACCTGTTG  
TTTCGTGACTTTTTAGTGATTGCCATTCTAACTGGCACCACAGTAAATTT  
TTATAGATTTTATAAGCAAATTGTATTTACTGTGCAAGAATTGGTTTATT  
TTTTAAACCATGTGTTGCAACATACAATGGTTAATTGTGATATTTGCTC  
AGTACAAGATCATCAGATCACTACACAGACTTGAGGTAATTCCACCTAAA

AGCAAAGAGAACTGACCCACATTAACTGAGAAGTCTTTACTTTATTAT  
CCCTATAAACGAGCCAATATGAAGAGAAGGCCTTAATGTGGTTAACTATG  
TAATTTTTTTCTGACTTTTTGAAATACTGAGAAGAGCTCATGACTCTCCC  
ATCTCCTAATTCTACCTTGGTGGATTTTAGACTGACCACAACCTCATGGGT  
AAATGAGGGAAGACGAATAAGAAACCTTGCTTTTTTTCTCCTCCTTGTTTT  
TGGCTGGCTGCAGTGGCTCACACCTGTAATCTCATCACTTTGGGAGGCCA  
AGGTGGGAAGATCACTTGAGCTCAGGATTTCAAACTGGCCTGGGCAACA  
TAGTGAGACCCCATCTCTAAAAAAGGCGACGG  
GCGGTGCGTGCCTGTAATCCTACCTACTCAAAAGCCGAGGTGGAAGAT  
CACTTGAGCATGGGAGGTCAAAGCTGCAGTGAACCTTGATTGCACCACTT  
CATTCCAGCCTGGGTGACAAAGCAGGACGCTGCCTCAAAAAACAAAAAC  
AAAACTTTAATTTTTGGCTATTCTTTCTGGTAAGAATGGTATAGAGAT  
GGGATGAGGATGGCTATTGTATGAGAGAGCAACAGGGTCCAAGCAGTG  
CTCTGGGCTGTCTAAGGACCAGTAGTCAGCTTAACTTCTCAAATTTCCAG  
GGAAGGAGTTCGGAGTGGTAGAATATCCTGGGTATGCCAAAGCATCACC  
TTGCAATAGCCTGTCTGAATAATTTGTTTCATTTGTTATGACTGGAAA  
CTGGCTTTGTGTATGCCAGAGAATGGGGCAGGAAAGAGAGATTGGTGTG  
TTGAGCTCTCTGTGCCTCTGGGGCAGTGATGCTTTCTCTCATGTGGAA  
GGAGAGCATGACTGAAAAGGTGCACAAATAAGGTGTCTGTGAGAGAAATT  
AACCTTCCAGATACAGAGACACAACCTTCCCCAAGAGGTCTCATTTGCTC  
TGCCTTTTTCTCTTTTTTTGCTTGTCTACCATTAAATAACAGAACTGA  
TTATGACCTCAAAAGAGAGAGGAGAAAGCGACTCTCCCCACCCTAGAGCTAG  
TTAACCACCATATCTTCTAGATCTCAGTTCAAGAGTCACTTCCATCCCC  
AATAAAAGCCCTTGAGTGTCTGAGCACCTCTCCGTCTAGCATTGTCTTA  
GGGTTTTTTGTACATTTTCTTGTGTGAAACTTGGGTTGACATCTGTATTT  
CCGACTAGATTACAGTTTCTCAAGGGTAGGGATGTCTTGCTTGCCATTT  
TCAGTTCCAGCATCTAGACAGTACCTCAAGCAAACAAGGCCGAGGGGGGT  
GCGGATCACGAGGTGAGGAGTTCGAGACCAGCCTGATGAACATGGTGAAA  
CCCCGTCTCTACTAAAAATATAAAATTAGCCAGGCGTGGTGGCAGGTGC  
CTGTAATTCAGCTACTCAGGAGTCTGAGGTAGGAGAATCGCTTGAACCC  
GGGAGGTGGAGTTGCAGTGACCTGAGATCCACTGCACTCCAGCTTGGGT  
GACAGAGCAAGACTTTCGTCTCAAAAAAAGAAAGAGAAA  
AGAACATCAAATGAATGAATGAGTGAGATGAATGAGTTAGCAGTGTGGA  
TTTAAGTGTGAGATTCTTCCAGCTTGACTTTTTCTTTGGCTTAGTGAT  
TTTGAGGTCNCAAGATTTATTTTCTTTTCAAAAGGTGATCACTACCATA  
AGATCTTCAGAAAAGAATGTGGCAAGCCANGTCTCACTAATGCAAATCT  
CTATAACAACCTGTATCAGTACT

>Contig43

GAGGTGTCATAAATATGGACCGATAGATGAATACAGGTAGGATGGGACAC  
AATCTAAGATCCCAGGGGGGGAGACCACACGCTTGGTTAGGGAGACCCA  
AAGTGGACCGTGTGGCCAGAAGAGTCCCGCACTGCACTCTAGTGACAGTG  
CAGAAAGTCACTGTGGGAAATCTAGAAGTTTCTACAGGTGCTATTTTCA  
CATAGCACTGTGCAGGCCAACCTTCTGCTCCACTGGCTGTTGGGAAAA  
GCTTTCTCTTTTCTTCTAGCCAGGGAGCTCTCAAAGTGTTCACCTCTCT  
CACCTCCACCCAGGCGTCCAGGTGTGGAGGACACTTGCCGGCTGCTTGTC  
TGCTGACTCATCCCTTGGTTTTCACTTGGAACCTACCACCAGCTGGCCT  
CTTTCCAAGCATCAGCCTCCTCATTTTCTTAATCCCTTAGGTGTGATCTC  
ACCTCCACACAGTAGATTGCCTCAAGGCCCAATTCCAATATGAATAAAAA  
TGATTATTTTGTGATCTTCCAATCTTCTTTTAAAAATATTATTTTATAAT  
TCCCTTTAGGAGGATCACCTAAGTGAAGACTATTTTTTACCTAAGAAATGT  
TAAAAATGTAAAGACATGGTTGTAATCTGGGGATTCTGTAAATGGCTA  
GCAGACAGAAGTCAGACGACAGGCTAGAAATGTGTGAAGAGTGGTTGCCT  
TTGAAAGGCGGAGTTGGTAATGATTTTCTTCCATTTTCCATGCTTTCCA  
ATTCTCTACAAAGGCCTTAATATTACTTCGATAACCAGGACCTCTGATAA  
CCTGCCCCCACCAGTAAAGACTTAGCTGGGAAAGTCAGCTTCATGTGAG  
GTAAAGGAACCAGGTAATACCAATTCCCACTGCCAACTGTGGGTGTG  
CAGGCTGAGCTTCTGTCATGTGGGAGGAAAGAGAAAGAAGAGAGAACT  
CCAAGATCCAAGAGATCCAGCAAGAAGGCTGGAGTCTGAGGACGCAGAAA  
GCTGAATGGCACAGTTACCCTATTGTGCTGAGGTTCTGTGGCCTCTGGG  
TCTCTTGACAACCTGGGCAAAGACCCACAGAAAATCTCTAGACCCTAC

FIG. 4 (21 of 61)

75/118

CTGTGGGAGGGGAAAAGTCTTCAGATCACTACAGGACAGCCACCTGGAACT  
CTCAATGGCTTACAGTTCCTTCATCCAGAGGGTCTTCATCTAGTACATA  
CCAGGTGCTTAAGCCTGGGTGCTGGAGACATGACGGGGAACCCATTTACCA  
TGGCTTTGTTACTGTGACATTCACATCTAGGGAAAGCCAGCAAAGGGGAG  
GGATCGAGGAGAGCTTGTAGGCAGAGAAAATACCCAAGGGCAAGGGAGA  
AGCCAGCCTGTTCTGAGCACACACAGTGGTTCATCTAACTGGGCCTCAG  
TGCCAGGTGGACTGGAGATGGGGCTGAGGAGCTGTCACAGAGCATTCTG  
GACACAGATGTCACATAGTCCCTTGAGGTAGGGTCTTAGGCATGGCAG  
CATTGCTTTGAGTTTTTCCTTTTGTAAATGTTGCCATTTCATGACAATGTGG  
AAGATGGGTCTTGCAGAGAAGGGCAGGGCTGTGAGACCAGTTAGGAGAC  
TAAGATGTGAGCCAAGGAAAATGAGGAACACCTGAACACTGGGGCAGGTG  
CAGGGCCCAGAGAGAAGCAGATGGCTTCCTGAGGTTTTAAGTAGGTAGAA  
TCAAGGCAGCTGGTAAAGATCTTTTATTACATATAAACTGGAATAAGCCA  
TCTGCTCCAAGACAAAAGAGTAGGCGGAAAACAATACAAGACAGAAATGG  
AATTAGAACAAACCTGGGAGGAATGTGGAATTAGAGTAGAGAGTCCAACA  
CTGGCTGCAATCATAAAAATGTAAAAACAACAAAATTTGCTAGGTGTGC  
TTACTTAGAAATAATTAGCTGTCATATTAAGTTCACCTGTGTTATGGCTT  
AAATGTGTCCCCAAAATGTGATGTGTTGGAACTTGATCCCCAATGCAA  
CAGAGTTGAGAGATGGGACCTTTAAAAGGTGATTAGGTCATAAGGGTTCT  
GCCCTCATAAATGAATTAATACTGTTATCATGAGAGTAGATTCCTGATAA  
AAGGATGATCTCTGCCTCCTCCCCACAGCCCTCTTGTGCATGCTTTCCTG  
CCTTTCCACCTTCTGCTATGGGATGACACAGCAAGAAGGCCCTCACCAGA  
TGCAGCTCCTTGATCTTGGACTTTCAGCCTCCAGAACTGTAAGCCAAAC  
AAATTTCTGTTTATTATAAAATTACCCAGTCTCAGGTATTCTGTTCTAGAA  
GCACAAAATGGACTAAGATCATTAGATTATCATTTTTTATCAGACTGTTG  
AAGTGAAAAATAAAAATCAAATAAAGAAATTAAGAGAGCTGCATGCAGCA  
GCTCATGCCATATAATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATTGC  
CTGAGCTCAGGAGTTTCAGACCAGCCTGGGCAACACGGTGAAACCCTGTT  
TCTACTAAAATACAAAAAACTAGGCCGGGCGCGGTGGCTCACGTCTGTAA  
TCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCATGAGGTCAGGAGATC  
GAGACCATCCTGGCTAACAAGGTGAAACCCCGTCTCTACTAAAAATACAA  
AAAAATTAGCCGGGCGCGGTGGCGGGCGCCTGTAGTCCAGCTACTCGG  
GAGGCTGAGGCAGGAGAATGGCGTGAAACCCGGGAAGCGGAGCTTGCAGT  
GAGCCGAGATTGCGCCACTGCAGTCCGCGAGTCCCGCCTGGGCGACAGAGC  
GAGACTCCGTCTCAAAAAAAAAAAAAAAAAACTAGCCAGGCATGGTGGTGT  
GTGCCATAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGA  
ACCCAGGAGGTGGAGGTTGCAGTGAGCTGAGATCATACCACTGCACTCCA  
ATCCAGCCTGGGTGACAAAGCAAGACTACATTTCAAAAAAAAAAAGAAAG  
AAAAAGAAAAAAGAAAAAGAAATTAAGAGAAGGGCAGGTATTAA  
CCCCAATATCCCACCATAGGGACACATTAAAGTTTGCTTGGCCACTCCC  
CTAGCATAAATATGGAATGTCTTCAAGGACCCTCTGTTGTAAATACAAG  
GCCCCTGTGGATTAAATACAACCTGCAGGCTTTGAGATCCCTACTCTGTT  
GCCATCTCTCATAGGATTTGCAGACCAAATCCAAATACTTAAATAGCAA  
CACTCACAAACATGCAATCAGAGCAGAAAAGAACTTCTAAAAGGCCCT  
GAACTACACTTTATGAGAGAAGACAATAGGGACCTGAGGGTGGTAGAAT  
TTTCTCTCTATGCATCTATGTTTCCAGGGCTCACTTTCTCAATAAACTCT  
TAAATTGCTTTTAAAGTAAGGGAACAAGCAAACATTACATTTAAGAGAAA  
TCAATTTTATAAAGAAGGGGGGATGTCCAGGGTACTTTGCTTCCATGTTT  
TGCTTCCATGAATTTGTGTTTAAACAGAAGATGCAGAAAAACACAAATTA  
TTGCAAAATCAAGGAAATCCACTCTAAACATCCCTTGGTTTCCCAGGCCA  
GTGTACAACTGAAAACACATATTGTGGCTAATTATGTGTACAAATTAG  
AATGACAAGGCAAGAAAAAATACTCTGATTAACTAATAGCAGCCAA  
CACAGACAGCCTGTGTAGCTCGACTCTGCTGGTTTATAAAAGGCAGAAGA  
AGCAAACGGCTTCTGTGACCGCAACAGGAAGGGCCTCTGCTCTTAATAAA  
TAAATAACATTTAAATTATTCTCCCCATTGCAAAGCATTTTCCAACCTC  
ATTATCTCATCTGACCAGGTATTATTGTATCTGACCAAGAACTTGTATAC  
NAAATAAAGAATAAAAAATAAATATGGGCCANGCACAGTGGCTCATGCTT  
GTAATCCCANCACTTTGGGAGGCCCAGGCCGGTGGATCACTTGAGGTCAG  
TAGTTTGAGACCAGTCTGGCCGACATGGCGAAACCCCGTCTCTACTAAAA  
ATACAAAAATTAGCCCGGCATGGTGGCACATGCCTGTAATCCCACTACT

FIG. 4 (22 of 61)

76/118

TGGGAGGCTGAGGCACGAJAATTGCTTGAACCTGAGAGGCGGAGGTTGCA  
GTGAGCCGAGACTGCGGCCATTGCCCTCCAGCCTGGGCGATGAGAGCGAA  
ACTTCATCGAAAAACAAAAACAAAAACAAAAACACCTTAGAAGA  
AGCGTTCCTCCTCTTGCTTTCTGAAGACACTCTACGCTGAAACAGTAACT  
TTCAATAAACCATCTCTCTCACCGCACTCTGCGACTTGCCCTTGAATTCC  
TTTGTGTGCAAGATCCAATAAGCCTCTCTTGCGGTCTGGATGAGAACCTT  
TTTGTGGAATACTCTGACACAACAATTGCAGAAAGAAAGTCTCACATG  
TATAAAATAAGCAAAAAGATTCTCTGGCATCTGAAGAAACAATTTCTTG  
TCAATATTAGTATCACTATAAGTGTAGAACAACCTGTTGTATGATGCTAC  
ATAAAGTATATGAATCTGAATACTGTTGGATACAAAGGGAGACTATNNAA  
TGTAATACGTGCCCCGAAATGACTACACTGTTGGTGATCTTTCTTTCAAG  
AAGCANAATATTGCCTCNAACATCCTGTACATGGTATAAAATTTTA

>Contig44

CCCAGCAAGAACACCAATACAACGGGGGGGGCGTTCTTTGTGAGGGGTGG  
GGAGGTCAATTTTTTGGAACCTGCAGCAGGTAAACACACAAAACCTTCCACA  
GCTGCTACCAGCTTTCCAGGAGAGCCTGTGTACCTGGAGAGGAGAAGGCA  
AGTGCTTCCGAACCTGACTTGATGTCTTAGATTCTGCAATGCGTAGTCTG  
TAGGGACAGGCTGTAGCTTATCTATAGGCTTGGGCTGGAGTCAGCAAGC  
ATCTGGGCTGGCAGAAGATAAAAGATGCAAAGGTGGAGGAAAGCATACTG  
GGTCTGGAAGACAGACTTGGTGGGTGGGTGGCTGCTACAACACCCTAGTT  
AGAGGTAGAGGGGTAAAGTCAGTGTGTCTTCTGCACAGGCCTCTTCCCCAC  
CTCATTCTTCATTTCCCATACAGCCTTGCTGAGTTATTCACAAACATCTG  
ATTCAACTGGAAGCTGGGTGAGGATGACCTAAAGGACTAGTGTGATGCC  
TGCCCAGGGGTGTGGGCCCATAGTCAGAGTCCAGAGCCTCCTCTCAGCTT  
TTAGCACATCTCACCCACATCCTGGGTCTTAATTAGCAATATGAAAGCA  
AGCCAAGTGACAAGATTTTGTCCCTGGGAAGTCCAGAAGCACTCCTTTTC  
TCATTTGTATAAGCATAATGATTTTGCTTACATAAATAATCATGAAAATTC  
AAATCCCTCTCAGAAATCAGGTCATAAAACCATGAAATGCAGCATGTGGG  
CAAGAATCACAGGGAAAGGTAGGTCTTGAAAAGAAAGGATGGCAGGGAG  
GAAGAAAGCAGGGTGCCAGGGGCCCTGGGCTGCTGTCCAAGTCAGGTGGC  
TCACCGTCTCTGAGAACATTTCACTTTCTGGTAAATGGGGCAGTTGGAGA  
TAGAAGGGTTGGGTGAATGCCAAGAGTGAGCACAGCTGAGGTCAGTGCTG  
TGCTTGCACTCCAGGCGGGAGTAGAAATCCTGGGCCCATCTTACCTCCGA  
CCTCATTTCTCTCTGTATAATGTGGGGGTGGGGGAAAGTTCTGGTCA  
TCAGCCCTAGCATTCATGGTTCAATTCCTCATCAGTGATGAAAATCAC  
CAAGCAAGAGCAACAGGATGGAGAATAACCGGATGGGTGCAATCGGAGGTG  
CTATTTTCAGGTGAGGTGGCCAGGGAAGGCCCTCTGAAAGGTGGCTTGAG  
CAGGTGGCTGAATGTACAGAAGCTGCCAATCATGAAAGATCTGGGGTACA  
GCATGCCAAGCAGAGGAAATGCGAGTGCAAAGGCCCGAGATTGGATGTG  
GGCTTAGCACAAATGTGGCATGGCAAGAAGGCCAGTGTGGCTGAAGCAGC  
ATGAACAATGGGTGGAGGGGCTGAGAGGACAGAGGAGCAGGAAAGAGCCA  
GGCTTGGGTAGGAGAGGTGTCAACTTGATATATGATGCAAAGCCCTTGGA  
GGTTCCCAACACAAAAGCAATGATCTAATATATGGTTTTAAAAATGCCA  
CTCTTGCCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG  
GCCGAGGCGGGTGGATCATGAGGTCAAGAGATCGAGACCATCCTGGCTAA  
CAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGCG  
GTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT  
GGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATTGCGCCACTG  
CAGTCCGCACTCCGGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAA  
AAAAAATGCCACTCTTGCTGTGAAAAATTGACCCTGGGGGA  
AGGAGGAGTAGAAATGTCAAAAGTGGAAGCAGACCACTCAGGAGGTCAGG  
GCAATGGACTGTGCAGGAGAGACTGACATCTTAGACTCGGGCAATAGGAG  
AGAAGGTGGTGAGGATTATATTCTGGGCATAAAGGCAACAGAACTAGCTG  
ATGGCGTCAACGTAGGAGATGAGGGAAGAAAGAAATCAAAGGGCATTCA  
TAGGTTAGGGTTGAGTAACTGGGGATATTTAACAGAAATGGAGAAGTC  
TGGGGAAGGGCAAGTATTGTGGGGCAGGGGTCAAAGTTCTGTATTTT  
GGCCAAGTTAATTAATTTTGTAGATACCTCTTAGGTGTCCAAGTGAAGAT  
GTCAAACAGTCAATTGAATACAAAATCTGAATCTTAGCCCAGGATGGTCT  
CACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGAGAGGATCACTTG  
AGGCCAGGAGTTTGTGATCAGCCTGGGCAATAGAGCAAGACCCTGTCTCC

FIG. 4 (23 of 61)

77/118

ACACACACACACACA. AAAAAGTCAACAGGCATGGTGGCACATGCC  
GTAGTCCCAGCTACTCAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAT  
GGTTCAAGGCTGCAGTGAGCTATAATCACATCACTCAATACTACACTCCA  
GCCTGGATGACAGAGAGAGACCTCATTATTAAAAATAAAATTTAAAAAAA  
TTAATTAAAAATAAATCCAAATCTTCTGAGATTATATTAGGAGTAA  
CTGTATGTAGAAGGCATATAATGCCATGGGTACATGATACCATCTAAT  
GAATGCCACTGGAAAAGAGAGAATAGCTAAAACTGAGCACTGGGCACAC  
CAGCACAGTGAGGTGGAAGGAAGAAATGGAGCTAACAAAGGAGACAAAA  
GAGGAGTAGCCAGTGAGAAGAGAGAAACATCTGGAGAGAAGAGAGAGCAG  
CAAAAGGTGGGTGAAGGAGAATGTGGTCCACCAGGCCCAACAATGCTGAG  
CAGTTGAGTAAGTGAGGACCTGGCCACTGAATTTGGCAAGAAAGAGGATG  
TCAGCGGCCCTAGAACAAAAGTGAAGAAGAGCTTGAGGACGGAAGCCTGA  
CAGGAGTGAAGTGAAGAGAGAATGAAAGGTGGAGACATGGAGCCAAGGAG  
CACTGAGACTCCCTTGAGTAGTTTTGCTGTAAAAATAAAAGTGAGTGCAGA  
GACGGGGCAGGGGACAGAGAAATGCAGGGGTAGCTGGAGGGAGCCACAG  
AATCAAAAGAGGGTTTTTGTGTTTAAAGATGGTAGTTGTCACATAGCACAT  
TAGTAAGTTCATGTGAATCACAACGTAGGTGAGACAGATCACTAATGCAG  
GAGTCAAATCCTTGACAGAGCCCCCAGAGGAGGTGATGAAGGGAAGTGATG  
GACATCATTGAGATGCAAGTAGGTTAGCAATTCCTGGGGTACAAATAGGA  
GGTGACTCCTTCTGATTGCTCCTGTTTTCTGAATGAGATAGCACATAAA  
GTCCACTCAGCCATGTTAGCTGTTGAAGTCCTTGTGGCTGTATGCCTGT  
ACAGACTGGGCTCTCCTCTCCAGCATTTCTCTCAGACTAAGCTGAGCTG  
CACTAGCCGCTGCCACATCCTCTTGGGGCCATCCTCTGCCACACTCCACA  
TATTGCTGTGGTTTGCTTGCAACCCCTGGAAGGTCTACTGGCTGCTCCT  
AGAAGAGTCTGGGCGGCATCTCCTTACTCGTTATCACATGGTGCTGT  
AAGCAGTGGCCACACACTTTAGCTGGTGGGATGGGCCATCACAGGCAGTA  
AATGCGAAAGACTGCTCAGATTTTAAAGCACCCATGAATCAGTAGAATGA  
GTTTGAATTTAGTGCATCAACACACATTAACAAAAACAGGCAC  
TAAAAAATTAGTTGAGTAGGATAAAGCCATAAAAGATATTAACACAAC  
CCAGATAGGAGGTGCAAAATTGTCCTTACATAAATCAGATGGAAAAAGTT  
GAAAGCAGATAAGATAAAATAGGTAAGCATGACATTTAAAGGTATTCAT  
GGGACGTGGTTACAAAACCAACTCACAACCTAAAAAGTCTTAGGACCTCTC  
GCTGACTTAGGAGCCTGATCCCAACTTTGAGAATGACTCAGTGTGTTACC  
CTGTGGCTAGTGAGACCAATGATCCTGTCTCAGAGTCACTAGCCAACAG  
CCCATATCAAGTAATTGAAACTTTGACTCAGAAACCTCAGTGTGAGAACC  
TTTGACTTAGGAACCACTGTAGTGGTTAACTGCAATTTGCACCCCTTAG  
TTCAGGGCTTTACAACACCGGGGGGGGAGGGGAAGGCATAGAGCTGA  
TGACCTAAAGGAAACCCATTGCAGCAACGCTTTTGTGTTAAGTTTACAAA  
TAAGTGTTGTTTTAGAATCCTCCAGGTAATGCCTTTGTTATTTAATGTGT  
CTGAGACAATTCGCACATTAAAGAATATAAAATATTACCTTGTAATTCC  
AATTTGAAATGTGTAATTGACATTAGACTTCTATTTTAAATTTGAAATGTC  
TAAACAATGTGGTTAAGTTTGTAAGGTGTGTGAATTTTGAGTCTGAT  
TTACTACATTTTTTTTTTAATTTTTTTTTTTTTTGGAGTTTTAGGGATTGC  
TTAGATGGCTAGAAAGATCGCTAGGCACATGTCC

>Contig45

GATGTGTGTACGTGTGTGCAAATACCGTGCCTTTTTTGTCTTTTGT  
GAAACAGAGTCTCACTCTGTGCGCCAGGCTAGAATGTAGTGGCGTGATGT  
CAGCTCACTGCAACCTCCGCCTCCAGGTTCCAGTGATTCTCCCGCCTCA  
GCCTCCCAAGTAACTGGGATTACAGGCGCCACCAACACGCCCAGCTAAT  
TTTTGTATTTTAGTAGAGACGGGTTTACCATGTTGGCCAGGCTGGTC  
TCTAATCCTGACCTCGAGATCCACCCACCTCGACCTCCCAAAGTGCTGG  
GATTACAGGCATGAGCCACCATGCCTGGCCAATACTGTGCCATTTTATTA  
TCAGGGACTTGAGCATCCATGGATTTTGGCATCCATAGGGGTCTGTAAAC  
CAATACTGCACAAATACCAAGGGACAACGTATTCTAAAAAGACCAAAAA  
TTAATAAGCAGGACGCTGAAGGTAATTGCCCCAATAAAGTCATGATCCCT  
TGCCAGTGTCTGAACCTCAGCCAGTTTCTAATCAGGACCTATTGGCT  
GCAGAGGTGGTAGGAACCATATGAGAATCCTGCAATATCATGGCAAGTAT  
GCACTTTAATGATATCTGCAGTCCTTCCCCAAAAGGACCTTACATTTACC  
ATACTGCTATGTCCTGCGTGAGAGGGTAATACTCAGATTTTTTTTTTTT  
TTTTTTTACACAACGTCTTACTGTGTTGCCACACTGGAGTGCAGTGGCT

CGATCTTAGCTCACTGC .CTTCTGTT1..TGGGCTCAAGTGATTCTC  
GCCTCAGTTTCCTGAGTAGCTGGGATTACAGGCGCCCGCCACCATGCCTG  
GCTAATTTTTGTATTTTTTAGTAGAGACGGAGTTTTGCCATGTTGGCCAGG  
CTGGTCTTGAACCTCTGACCTCATGTGATCCGCTGGCCTCCCAAAGTGCT  
GAGATTCCAGCGTGCGCGGCCATACCCGGCCGGGAATTCTTTATATATTC  
TGAAAACTAATCCTTTGTGAGACATAAGTGTGTAAATATTGTATCCAG  
TTTGTGGCATGTATTTTTAATTTTTAATGGTGTCTCTCAATGAAAAAAGC  
TTAACACTTAAATGAGGTCAAATTGATCACCTTTTTATTTATGGTTGATT  
CCTTTGGTGTCTATGTGTAAGGAATGTTGTTCTTCTGTCCCAAAGTTGC  
AAAGATTTCTTGTTGATTTTTGTCTCTAAAAGTTTTAAAGTTTTGCTTTTCC  
CATCTGTGCACATTTACATTTGCTACATCTCACTGACTGCTTCCTCTGC  
TGCAGAGCAAGCTCCATGAGAGCAGGAGGCATGGGTCTGTCTTCTTGTG  
GTCCCCAGAGCCCTATGTCATGACTAGGACCTGCCAGGGGACTAGTGAGT  
AGCTCCTGACTAACTGACTCAATGAATGAATGATTGGATGATTGAACAAA  
GTGGTATGGGAGTTCACAGCGAGTAAGAGATGCCTTAGAAGAGATGAAGA  
AGGAGATGGTATAGGGTAGTGGTTCTCAATTCTGGGTCCATGGTGGACTC  
ACCTGGGGACCCTTAAAATGTACCGTGGAGGATCCCAGCCCAAGAGATT  
TGTATGACTGGTCTAAGATGTGGTCTGGGCACCAGGTGATCCCAGTGTGC  
AGCCAGGCCTGAGGCCACTGGATTTGGTGGTAAATGAGGTAACTATCAAG  
GGTACAGACGTTGGTTGCCAACAGGCTTGGGCTTGAATTTAAGCTTTGTC  
ACTGACTTGTCTGTCTCTCTCGCACTCGTTGAGCCTGTTTTCTCAGCTGA  
GAGATGGGTGTGATAACACCTACCTGCTGTAGTTGTTGTGAGAGTTAGAG  
GAGATAAGCATGTTCTCTGGAATGAAGTGTGTTCTTAATCCATCATAGGTT  
TTTTGCTTGTGTTGTTGTTGTTGTTGTTGTTTCTTTTCAAGAATGA  
GGTTGAGCCAGACTTTGACAGCTGGGTGGGAAGTGAACATGTGGTGATTG  
GGAGAGAAGGGCAGTTTATGTGAAGGGAATGTAATAATTAGAGAGTGGGC  
GTGGGAAGACATGCTGGGGAGAGTGAGCAGGCCGGTTAGCCCTGGTAGAG  
GGTGCAAGAGAGCAGTGCGGAATCTGCCAGGGAGACAGGTGGGTGACCAG  
GGTGCCAAGGGTGTGGCTTTTCCCAGGTTCCCATGGACACAGCCATCCTC  
CCAGATGCCCAGCCTAGCTGTGAGTGAGCAAGAGTTCTGGATTGTCTCTC  
TCACTCTGTCTTTTCTCTCATTCCAGAAACAAAGCAGTGACTGGTACTT  
AGGAGGAGAAATCAGGTCAAGTTGGGAGAACTTGCTTCTGCTCAGGGGAG  
CAGAAGCAAGAATGGAGGCCCCACCCATGCTGGAAGATGATGAGGGTTTT  
GGTTCAGGGAGGAGGAATATTGGGGATCTAAAGGGGCCTGGGAGTGGGGC  
AGGACCCTGCCTTAGGACAGGTAGAAACATTTTCTATAAAAAATGGGGTG  
GAGGTTGATGGTAGGACCAGGCATCTTAGTTGGCTCCCTGGAGTGTCAA  
GCCCTTGAGATGGTCTTTAAAGCCATGCAGTGGGGTTTTGAATCTGGTGT  
TCAAGCTCATAGGTTATTAACATAATGACACTTGGAACCTATTGGGAGA  
GCTCAAGTGAGTGGCCTGGAAGTTCTGTGTTGGTGCAGGAGGTGACTTAG  
GATGTGCTGCTCCAGACTCATATCTTTGACTGCACACCTGATGCTTCATC  
TGGCTATCTGTGTAAGCACCTTCAACTTAACATGTCCTACACAGAACTCTT  
GATATTCTGTCTCTCCCCAGTTCTCTCAGTTCTTACCAAATGTTCTTCC  
AGTTACCCAATTGCTCAAGTAAAAAATCTAAGTCCTTCTCTTGGATTTCT  
GCCTGTTCCCTCAACATCCCACCTATCCATGAGTGTCTGTGGGCCCTGC  
CTCTGAAATAAATCCTGCCTTTGTCTCCAGTTCACTCCAGCCACCCATC  
CTGGGGCTGCACCCTCCTCCTTCCAAGCCCTCTCCCTTTCTTCTGCTG  
CTGCCTGTCAATGCAAGCATATGCATCAGTGCGACCAGGACATTTGAAAT  
GCAACCAGTACAATTGGGCGCGGTTATGCCTACCAGTTTTTCTTCTCTTAA  
ACATTTTATATTTATGTTTGAAAGCATGCCACCTTTCTTCACTTGCCAAC  
TTGACAGATTTATAGTTGACAACATCCGCTGATAGCATCAGTAATAAGT  
TAATTGTTTTTGCACATGTAGCTTTAATTATTCTCATTATCATTATAGG  
AGTTATTCTTTGTAAAGGGTAACTGAGTTTTCCAAAACAAACAGAAATTT  
GGGGTGGGCCCATGGAGCGTGACTCATGAAATCAGATTCTTAGAAGGACC  
TCGGCAAGTCTCTGGGTTGCTGTTAATGAGCCTGGCTGGCTGCCAGGGGT  
GTGTCTGCCCTTTATGAGGCCACCACTGTTCAAATGCTTGCCTGCAGCAT  
TACTTGCTAGGTAGTGCTTGTCTTCTACTGAACTGTCAGGGATCCAATTC  
TTTGTGGTCTAAGTAACAATACTCAGATTCAAGGAATTGATTAATAAG  
CCAGAATGCCAATGTATTACATTTTTGATGAAGACCATATTTACAGTGAT  
TGTATCTGCTCAAGCTCAAATTAGGATTAGAGTTCTGACAAATACATATG  
TGAGAAGTATGAGGTTAAATACTTGAAATTTGGACTTTTCTAGAAAATCT

FIG. 4 (25 of 61)

79/118

GAATGTGATTGCCATTACATACCTTTCTGGGGATGATGATTCTTGTACT  
TTTATTTTAAAGACATAGAAAATACTTAAGAATCAGATTGCTTGGCT  
GGGCACAGTGGCTCATGCCTGTAATGCCAGCACTTTGGGAGGCCAAGGTG  
AGTGGATTGCTTGAGCTCAGGAGTTTGAGATCAGCCTGGGCAACATGGTG  
AAATCCCATCTCTACCAAAAATACAAAAAACAACCAAAA  
AGAATAAATTAGCTAGGTGTGATGGTGCCTGCTTGTAGTTCAGCTACTT  
GGGAGGATGAGGTGGAAGAATTGCTTGAGCCCAGGAGGTGGAGGTTTCAG  
TGAGCTGGGGTTGCAACAGTGTACTCCAGCCTGGGCGATAGAGTGAGACT  
CCGTCTCAAAAAAATAATCAGATTGCTTTATTGCTGGTTTCTTTCT  
AAAATGAGATTGGGTCCCATCATCCCCTGGCCCCCATTGGTTAATGGTT  
CCTCCTTTGTCTATTGAATAAATAACAGATGTCTGCTTTTGGCAACATGG  
TTGAATGTAGACACTGCAGGGTCTTCTGACTCAAAATGATTTAGGCTTA  
GATAAAACACATTTGGAAATGCATTTCTGGATTAACACCAAGGAAAGGAG  
ATCTCTTTAAATCCCCCTTTCTGTTCCCCCTCCCTACCCCTCCAATTGG  
GCTTAAGTAAGAAGGGTGGTTACCCGCTAGTAAACCCCTTCGAAGGGGG  
TCTTCTCCTCTAAGGGAAACCCTTGTTTTGACATTTGCTTCAATGGGCC  
CTTGATTTTGTTCCTTGCTAAACGGGTGCTAAACCAGGGGCCTCCTCTT

>Contig46

AAGGCTTTTAGAATATTTGCACACTTTAGAAATGGAAATGTTTTTGGGGG  
GCGAGTTGTCTTAATATTTTCTAGCTTGTGTGACATCCTTTTGA  
AAGCAGCAATTCTGGCCTTTGTGAGAGATGGTGAATGCCTGCAGGTGTGT  
GGACCAGTGCCTCCCTTCCTTCTACATGCACGGCCCCCAGCTGGGCCCCA  
GCAGAGTGCTGTTACAGAATAATTTCCAAGGGCTGTGTCTTAACCTTTG  
GTCTTGTCCCCATTGCTGTAGATTTGGCCAATTGACTTCATAAGTGCCT  
CTTATGAACATAGATGTTGGCAATGGAAGTTGAGGACCAGTCAGTGGTTG  
TTTTATTGAACACACAGCGTAAATCCCAACACAATGCTGACCTAAGAGAA  
TTCCAGCCACTCTGATTCTCAGTCTCTTTATATCTGAAAGGGTTCTGTTT  
CACTTTTCCAGATCAAAATGTCCCTGCAGCTACTCAGCAGAGCTGTCTG  
CAACTTATACGTAGAAGAGGTAAACAGTCCACAAACAGAAAGGCACAGGAC  
GAGAGTGGTCTGGGTGATGCTTCCTGTGGGGGAAAAGGTGATGAGGGTGC  
ATCTGCACACCTATGTTCATAGGTAAGTCTGGGAGGAGGTGACCTCCCCCT  
TTGGTTGAGGTGCTGAGGCGTCTTGTTAGAATGGCACTATTCCATTTATC  
TGATGCAGTCTGTGGGAATTTTGTGGTATGGCCACCACAGGTACCATGCT  
GGGAACAATGCCAGATACTGCCTGCTAAGCCACAGCATGAGTCACATGAG  
CATTTGTGGGCTTTGGGAATAAGTTATTGAACGATAGTTATCTGAAAA  
GGAATTTAGGGAAAGGGGACTTTAGTCCAGCGAACAGTTTGCAAACCAGG  
GGGAAGGCAGCCTTCAGCGTAAATGAAGACGTGTGTGCCCCAAATAACA  
AAGGGAGAGTTTGTCTTTTAGAGAGTAAATGTCCACGCAAGGTTCCACTT  
AGGCAAAATGAAAGATGCAAACTTGCTTAGTTCTGATTTGTTTACATTTGC  
TGAATTCGGATTGGTCCGTGCAGGCTTTTCTGGGAATCCAAATACATGT  
ATGACCTCTAGTCATACATGGCAAATGGCCGCTTGCTCTAATTTGAATT  
TAGGCCCAGTTAGTCACTCAGGATTAACCTTTTTCAGGGTTACAGCTCT  
GAACAATGGACTTAGACCTGCAGGACATAATCTGTTCTTAACCTCTGGGAC  
TACCTGTGCCTTTTGAATGTGCCAGTGAGCAGCTGTGGCTCTGGGCCCCA  
GACCCACAGGGCGATAAGGCACAGAGGTACGCATGGAGCAGGCTGTCTT  
GCTGAGTGATCATGAAGATACACTTACATAGAGCAGCACTTTTCTCTCCA  
GTCTTTGTGATTTAACTCATTAGATCCTTATAACAAGAGTCAGTCCTCTA  
TTTAACCCATGAAGCACAGGTGGAGTCCAAGCTTAGTTTGTGAAGGATGA  
GCCAAAAGGATTCTTCTCTGTAGACCTCAAGCTCAGCTCTCTCCATGGG  
CCCTGGAGTAGGTGAGAAGGCCTCTGTCTTCCAGAGCCCACTGCCAATCA  
TCTACATTTTCTGTTAGCCCAATTCTAGGACATTGCTTTACCAACTGAAG  
GGTGAGAACTATCATAAGTTATAAAATCAATTGAAAAACAAAAGGTAC  
AGAACAGAAAATAAAAGATGAGAATCTATTAAACATAGTGATGTTACTGG  
AAAAGGGGGTCTCAAACCAGACCCCAAGAGAGAGTCCTTGGATTTCACAC  
AGGAAAGAACTCAAGGTGAGTTGCAGGGTGCCTGGAATTGAGAGAGTTTA  
TTGAAAGCTATTCCATTACAAAGTAGAGCATCCTCAGACAGCAAGTGGAG  
GAACATGCCATCATTAAATTTTCTTATATAGGAATCTTGTCTATATAAA  
GACTAAACTAAGCTGTGGCTATGTGTGGGTGGGCCGACAGCATGAAAACA  
TTTATTCTCCTATTGATTTAAAGAGAACTATCCTTGACATTTTAGTGTGT

TTAAGTACATCAAAGCAAACTATAATTAICTTGAAAGCATATATTTTTTA  
TAGGGATTGGGACATCTGGGCTTTCTGTTGTTGTAGAAGTTTGTCTTGC  
AGGGATTACCAAGCCACTTCCTTAGCTGTAAACATCTTAGGGCCATGGGT  
CCTGACTGGCAAGGAATGTGTCTTGCTAGTTTTAAGATGGGCTTGATTTG  
AAAATGGTGTCCATCTGGCTCTCCTAGGCTCCTGCTTTCCTAACAGTAAG  
GGTAAATGCTATGTTATGAAATGTCATTTCTGCCTTTAGCTTGCAAACCTC  
TTGATGGTGAAATTCCTGTCCGTTTTTCAGTGGGGTATTTATTCTGCAT  
CCACGCTTTCACAAGGAGCTGAAAACAAATTGGATGGAAGCAACTGGGTT  
TTATGGGACACGTTAATGTTTTAATGTCATTTGGTGTGGAATTCAGATGT  
CCAAGCAACATTTTACACTACAAATCTGCAACTTTAATAATCACTCAAAG  
TACCTGAACCTCAATGCTTTCAGACAGACTTGGTATAAAGCCACCACCTC  
TTTCTATTATGGCAGCCCTATCCTGAGGACACAAATTTCTGCAGGGCTTC  
TGGCATATCTCTGATTAAACAAATGTCAACAAGGTTAAAACAAATGTCAT  
CTCTGATTTGTTGTTTTAAAGCCTGGATTTACTCATTGAATATTTCACT  
CCTACTAGCATGTCTTGATAGTATTTTCTTCAGGGACCCTAATTATTGCT  
ATTAATAATATGTGTGCAGCTACATGTTTTTTTTTTTATCAATTTGCAATG  
AAAACCTTTAATTGAATAATCTATTAGTGTTATTATTTGAAAGTGAAATCT  
TTTTCTTTTGCTTTTCTTGTCTCACACATAGTGCAGACAGTTTCCACACG  
GGCTCATAAAAGGAATGATTCTGCCTTGTGTGAACCTTTTTGCCTTTATTG  
TTAATTGCACCATTTTGTGACTGGCTTCTTGACCCTGTTGTAACCAAGCT  
CATAATGTACATTATTTCTTATTTTGCAGTTGTAGACACTTGAGGAAGTT  
CCCATTCTTTGTTTCTTCTTGCTTTTGTTCCTGTGATAACTTTTTCATG  
CAGACATTTTTTTTTTTTTTTTTTTTGGAGACCGAGTCTTGCTCTGTCTC  
CAGGCTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTCTGCCTCC  
CAGGTTCAAGAGATTCTCCTGCTTCAGCCTTCTAGTAGCTAGGATTGCA  
GGCGTGCCTACCAACCCAGCTAAATTTTCAAATTAGCCACCCACCT  
GGCTAATTTTTGTATTTTGTAGTAGAGACAGGGTTTCAACCATGTTGGCCA  
GGCTGGTCTCGACCAGGTGATCCACCCGCCTTAGCCTCGCATAGTTGCAG  
GTGCTATTCTGAGCTCAGGGCTCTGGCAGCTACAAGCCCAAGATGCGGTC  
TCCAACATGTGGCCATTCAATGTCTATGGCGCCCTCTACTGGTCTGGGAA  
GCGCAGCTCTGCCAGTAGCTCCAGCAGGGCACAGCTGTTAAGTCGTGATG  
TTCTACAGGTGACCAAGGGCAATCTCTGGACTCCTTAGCCGCTAGGTCC  
TCTCTGTAGCAGGACCCAGGAGAAGGCAGGGGCTGAGGATGGCTCTCTTA  
GACATTTGTGATGAACCAACGTTGTGCATTTCATGAACTTCTGTGAGCAA  
GCAGGTGAGTAGAGTTGGGTTATAAAAAGTCTTAGGGTCTCACTACAGAG  
ATGGACTTGTGTTGTAGATGGTGCAGAGCCGCTGAAGAGTTCTACTTGGG  
GTAATGGTGTGATTGGGTTTTCGTTTTAGGAAGATTTCTTGGCCAGAATG  
AGGCGGGCAACCCAGAGCAGGGAGTGGCCACATGTGGGTGTGCAGTTATG  
GGCCACTAATCCAGGTGATAAATGGTGTCTCTGAACTTCAGGTGGGGGTG  
CCACATGTCTCCATCTGCTCTGTACCCTTGAGACTGGCCTTATGGGCTGC  
CTTAGTGGTCTGTTGTCTCTATCTCCTGGTTGGGCTCAGGCAATGGGAG  
ATCAGAGGGAGGAAAGAGAGCTTGGTTAGAGTGCACCCGCGCCCTTCAG  
GTTGGCAGTGGCCACATTCCCCTATACAGAAGGCCACAGTTTCTGTGAGT  
GGCCCTCCACAGCCCCAGCTTCTCAGTGGGCCAGCCACCTCCCCATCC  
CTTGCTCCTCCTCCTCCAGAGAGGGTTGTGGATTTCCACTGTGAGCAGTG  
CCTGGAGCTCCACCATCTCCTGCTGCTTCTCCTGGACCTGCCTGCAGTTT  
TATAAATAACCTTTCTTACATTACCTCTAGCATGCACCTTTTGTGTGTA  
TACTCTGCCCCCTGTGAGCAGTACTCATGCCAAAGAGTTTGAATTTTT  
TTCTCCAGGCAACGGGAGGTCAATGGAGGATTTTAGACATTGAGAACAGA  
TGTGTATTGTGGAAATATCTGTCTGACTGAAGTGACCAGGATGGTCCAAA  
AGAGCGAGAATTTGAGGCAAGCAAACCATCAGCAGGCCAGCAGCAGAAAT  
CCAGGTCATAAACAGGGAAGCTGAGGCTCACAGGGTTGGATCAGGGAATG  
GGAGAGGGAAGCCAAACAATTCCATGAGCATGTGAGTTGCACATATGACT  
TGGTAATATTTTTATTTTTATTTTTATGTTTTGAGACAGAGTCTCGCTC  
TGTCACACAGGCCAGAGTGTAGTGGCATGATCACAGCTCTCTGCAACCTC  
TGCCCTCCTAGGTTCAAACAATTCTCCTGCCTCAACCTTCCAGGTAGCTGG  
GACTACAGGTGCGCACCACTACACCCAACTAAGTTGTGTATTTTTAGTAG  
AGATGAGCATTACGCTGTTGCCTTAGACACGG  
>Contig47  
AATATTGATTATTTGACCAGAAATTCATGCAGCTAACCGTGACCCCTGGC

FIG. 4 (27 of 61)

81/118

AAAATAAAATAGTGTAT...GTACGTGCAATATACATGCAAAGAAATGAGI .  
GAAACTAGAAGGATGTCAATCAAATGATAACATGGTCATCTTGGGGTCGG  
AGTACATTTGGGGATGAGGGGAGCTGTAAAAGCAGACTTGGACCTTTTCT  
TCTACCAGTACCGTGTCAATTTGAATTTTGGAAAGAAAAAACTCAG  
AAGGAGGAAAGCAGGAGGAGAGAAGAAGATGGATCTTAAGTGATTTGC  
CCGGGAGCACCTTGAGAAGGTGAGATTCAAGTCTAGGTCTAAGCTTTCTA  
ATTCCATGAGTGGGAGTGACCCACGTCCAAGAGGAAGCTCAAAGGAAGA  
TGTTCTCCATCATCTCTTGCTCATCCTAACAGCATGCAAAACCACATCCA  
ATGCAGCTCAGAAAACCTCCCAAATTGCCAAATTTCAATGGAAACACTTAA  
TGCTGTGGTTTCCAATTTCAACTGTAAAGTAGGTATGTATGCCATTGTTA  
CCATTAACCTTCTCAGAAATGGAGAGAGCTCTCTTCCGCCTCCTCCCCCT  
CTGCTGTGGCTTTGGTGAGACGTGCACTCAGGCTCACCTGTCTCCATGAT  
CTCCAGTAAGTACACATGAGCAGAGAGGCTCAGCTCAGCTCTTCTGGT  
CCCACAGGGTTGATTCTTTGAGAATTCTAGAATGCCACATCCTAGGCCC  
CCCCAAGAAATCCTGCATCTTACCCCCAGAAATATGAATCATAGCAAAT  
TCAAATCAACCATCGTTTAATACTCACAGACTGGGCACATCCAAAAACAT  
ATTTTCAGTTTTACAACAGTGCCTGGTGCAATCGGCACATTTGTGGAA  
GCAATAAATCGACACGGAGCTGAAACACAAACAAATGCCAAATTGTTTT  
ATAACACCTGATTTTCTTTCTGTTTCTTTATGCAGTTTAGTTTTGTTTTG  
CTTAACCTCTACCTCAGACCATAGTCTGGTAAACTCACCACCCAGAAGCTC  
CCTTGAAATGTGGGTATGCAGCCACTAGGTGGCAGGAGAGAGTTTCTGTC  
CTGGAGGGAGGACAGCCACTCTGTCCCCGGGTGAGGCCAGGGCCACCCTG  
CTACCTGCAAAATTAGCATGGGGCTTTATGAACCACAGCTTCTTAATAAA  
CACAGGATCTGTTTGATAGAGACTCCAAAACACGCCTACCTAGTGATGAA  
AGACTCAACTTCAGAAGAAAACCTTCATGGCAAACATCTTCAGAGATGTT  
TCCAACCTAAGGTTCTGAACACAGACGCTTCCCCAGAAAGCCATTGTTTC  
TCAGCACCTGGGAGCCTTGCTTTGCTTTGCTTACAGACTCGCTGTTCTTA  
AATCACTGCCAAGATAACATCTGTCTCTTCTTACCCTCTATTTGATA  
TAAGGACTCCTCACTCTTGTTGCTTCTTATTGGCTACCTCTCCACAGGGA  
GAAATCGCTGATTTAACAGCAGTCAATATCCCAAATCTGGAACAGGGAAC  
AGGGAAGCATTTAAAAATTGGAGAATTTAGGCCGGGCACAGTGGCTCATG  
CCTGTAATCTCAGCACTTTGGGAGGTGACGTGGATGGATCACTTAGGAG  
TTCGAGACCAAGCCTGGGCAACATGGCGAAACCTCATCTCTACAAAAAA  
AAAAAAACCTGTGAGCCCCAGCTACTCAGGAGGCTGAGGTGGCAAGACTG  
GTGCACACCTGTGAGCCCCAGCTACTCAGGAGGCTGAGGTGGCAAGACTG  
CTTGAGCCCTGAGGTGAGGCTGCAGTGAGCCGAGATCACACCACTGCAC  
TTCAGCCTGGGCAACAGAGTGAGACCTTGTCCAGATAAATAAATTAAAT  
TAATTTAATTAGAGGATTTAAGGATTTTCCCTACAGACACCTCCTTATTT  
TCTCTGGCCTTTTCTGACTACTCTCCCTAACTCCCTGCTCCTCTGGTCTC  
CCAAAACCTACTCCAGAAAAAAAAGGGGGGGAGGGACTAAAGGAAAGCC  
AGGTGACAGTGCCAGTGTGACAGATGACAAAGCATCTGCCCCGAACAAACC  
GTAGGTCCCTGAACTTTCTCCAAGACCTGTCTGTGGACTTACCTATGAAA  
ACCAGTTTTAGCAAAAACCTCCTAAGCCAGTTTATCAAGATCCCCTTAT  
CCTCAATATCCATCTGATTGGATTCTTCATCCCCCACCATTCCCCAGTGA  
TGTCACCAGGCCTTTCTTCAGCAACAGTAGTTAGTGGGTGTAGCCAGGAC  
GCCCCCTCACCCCTGATATGCCCTTTTAGTAATTCTTCATCCACAGGTTT  
CCACCCTGCTCCTAGGCTATACATTCCCATTTGCCCATGCTGCATTCCGA  
ATTGAGCCCAGTTCTATACTGAGGTCTTACTTCACCTCTCGCCATAGTCC  
TGAATAAAATTTGGTTTTTACATTTTAAAACTGTCCAGCTCTGGTTGTTCC  
TTGACACAGGGTAATTTTTATTCCATGTGATAGTTTGCCTTACCTCAGCC  
TACACCCCTCAAACCTGCAACTCTATATTCAAGAACCAGACAGCCCTTTC  
CAACAGATAGGAAGAGGCTGCCCTGGTGCAAAGGAAGAGGCTCTGGGAGG  
AAGGAGAGAACCAGAGGCTGCCCTCCTCTAGACTGAGCTCTGGGATG  
GGTGGACGATAAAACCCAGATACGTTTAGACATCTGAGCGTGAGAGGAC  
TTTGCTTTGCTTCCACAGGGACCCCAAGGAACTGCAAGCCCTCCAGAGA  
CTAAAAACAGCAGAACAGCAAGAAATGGCAGCAAAGGTCTGGGCAGAATC  
ATCCTATGTGGGCACAGACACAAACAGAGTCCCCTGTGGCCCCAGGAGAG  
TTTAAAGAAGATCCAGAGGCTGTCTATTCCATATCTCAGCAGAGACAGG  
CCCGTGAGCCTAAAAGCTGATCATTAGGACAAGAAGGACACGAACTGTCC  
TGCAGCGTGAACCGCTGGAACAAGGCCAATCACCAGACACCAGACCAGC

FIG. 4 (28 of 61)

82/118

CAGACACAGCCCCGAGTCCCCAAGACCACCACGGACCCATCGCCCCCTC  
ACCAATAGCTCCAGGCTACATAGACCCCTCCACTTCATGGATGTCCTCA  
GAGCAGAAAGGGGAGGCAGGAGTGGAAACCCTGACTTGGTTGAGTTGAAAC  
ATAAAATGACTGTACTATTATTGAATTGCTGAAGTTTACGTGAAAGAAAT  
GAGATTTAGTTTTTGGCCACAGTGCAAAATAAGAAACGAGGCTTCAACTG  
AGATTAAGGTGAGTTATAGGAAAATGTACTCCCTTGAAGGACCTGTGAAG  
TGTGTTTCGCTATGAGAAAATGACCAGAATCCACGTTCTTAGCTGCGGGAC  
TCAGGCTGACTCCTGTTTCTGGAGCTTGACAAAGGGCAGGGAAATCCCT  
GTTTCAGGCACAGTGATTTCAATGTTTAAAAGAAAACAGGTGGGCCCTGG  
CAATCATGATAACATGTCTAAGTTTACATCTCTGTGAGGCAGGTAGTGT  
AATCCCCATTTTGCAAAGGAGGAAACCGAGGCTGAAAGCAGCTACATGGT  
CTCTTCAATGTGGCCCAAATGTTGGAGAACAGAGCTTAACTGAATCAGCA  
ATTCTATACTTAGAACTGACTCTCTCTTTATTATATCTCACTACTACCTT  
GATATTTGAAATATTCAACTTTTTTCAATCAAAAAATAACAATAATTTAG  
GCATAATGACTACTATGTCATTTAATTTCTTGCTGATATTTCAATATCCC  
ATGCCAGGAATATTGAAAGCTCAGCTCCTTAAGAGCTGACTATGGCATCA  
ACTCCCAACAACCATCCTTCCAGAAATATTTTCCCTTTCTTTTGTATA  
GAGTGGCACTGGCCTATATGGTGACCACTTGCCACATGTGGCTGTTGAAC  
ACTTGAAATTTGGCTTGTGAGAATTGCAGTGTAAAGTGTAAACACATACC  
AAATTTCAAAGACATGGCACATAATAAAAAATGTAAATATCTCATTAAC  
AATTTTTATATTGACTGTGTAAGTAACATTTTGAATATATTGGATTAAAT  
ACATGGATGATGCCCCAACACCCACAGTCCCTTATCAAGTCTCTACTTCA  
CATTTTTGTACTTCTGACTTAGAAATAGCACTGGCGTCTAAGAGCCTATT  
AATGTCGTCAATAGGTTCTTGGGAACCAATTTTAAACAAAATGACATA  
TAAGAAAACGAATAACATTGAACAAAATGACATTATTCGAGGACCTGCTG  
CATGTTGTTTCACTTAAAGTCAGTGTCCAAGAACCTATCAGTGACATTTA  
GTGAGGACTTGCTGTCTTCTGTTTACAGGAACCTGGGCAAGTTACTTA  
ATTCTCTAAGCCTGGTTTATATCCCTGCAAAGAGAGAAGGATAATAATC  
ACCAGTACTTAGTGATGTGCTAAGGAGAAAAATAAAATAATAAATATGAAA  
TGGCTGACAGTGTCTTGTGACACAGAAGATGTGTGATCCACAGTAGCTG  
CTATTGTCTGCCTCACTTCACTAGTAATGGTCCAGGGAGGCCTTTAATGT  
GCATGGTGCAGTACATTACATGTTGGACATGGGTGAAGGGAAAGACCAG  
GCTCATCTAAACACAATAGGATGCTTGTGGTGTTTTGAGGAGGAATCAAG  
GACTAGTTATCCACAGCTGTAACATGCATGGATCAAAAGAGATAAGGCAC  
ACAAAAGACTTTGTGAGTAGCAAAGCATTACAAAATGCAGAGACCAGCTG  
TGGGTGGTGGTGAGTCAGACCCAGCTTCCCTCTGTGCTGGCTGAGTGGT  
TCTGGGCAAGTCACGCCATCTGTCTTGATGCCCTTCCCATCTATAGAGA  
GGGAGCAACTGAGGCCCCCTTCCAATACTGAAGTCCTTTATTTCTGCTACT  
TTAGAAATATCCACATTTTTTGGTAAATTCAAATGATCCAATGATTCCATT  
TCCTAATGTTCAAACTAGCCCCAGAAACATCTAAATGAATCAAACAAT  
AAAATATTTATTGTGTATGTTTTGATTGCTGAACTTCTATTTTAGCAAC  
ACACACACACACACAGAACCCATAAGCCTTCATCTTTCCTTGGATAAA  
CGAGCCTTCTGTCTGGCCATTTAAGTCACGATTAAGTAAATGATTTCCA  
ACTCGCCTTTTGCAGCAGTTCAGATGGGTCTTTCCTGCGTGGCAGTGGCC  
CTCCTGACTTATGATTTCTGTGTGTGCGCCTGTTACCACTGCAGCTTAA  
CTGAGGAAACAAGAACAAACAGCTTCTGACCCCAAGAGACTGTTGGAGG  
CAAAGGCTTCAGTCCCAAGAACCTCACACGTGGGGAGCCCGAGAGCCAG  
CCCTGACCTTTTCTCCAGTAATAACATAAGAAACAACAGGCACTGGCCTT  
ATTTTGGATACAAAGAGTGGTGCTTTTCTTAAATCTTCTTTTAGTCAGG  
GCTACCCCTTCATGGACGCCCCAACATCCATGGTTCTGCTTGTGAGTCCCT  
GCTTCCATATTCTGCACTTCTCACTTGAAATATCCCTGGAGTACGTAA  
GCAGCCAGGTTTGAAGTTCTTGCTGTGACAGGCGGGTGTGTGCATGTCCT  
CTCTCTCAACAGGACACAAGCTCCCCAAATCAGACGGTATGCCTCCACGC  
CCCTTCCCAAGCCTCCCCAGCAGCACCGAGCATGTGAGGGGAGCTGGGGC  
CCAGGCCATGATGGGAAGCACTCTCTGCCTAAAGACTAGGGTGATGCGCC  
CTCAACTGTGGGAATGAGCCCCAGCTCTGGTGTCTGCCTCGGTTTTTCTCT  
CTTGGACATGAACATGAACCTCCTCACCCTCTTATCCACTTTGCATAAA  
CTGAAAATAACAAACCCAGGGCTCTTCTGTGACAGGAAAGGGTTTTTTT  
TTATAAAATTAACAGAGATGATTCAACACACCCAGGATATAACACATGG  
GCCATGAATCAAGGGCAGCATTGCTCTGGTCAGCCTGTTGTTTGGGCCCC

FIG. 4 (29 of 61)

83/118

CTTGGCAGGGCTCTCCCCA GAATCTTCCCCTCTTGACTCCCATCANCACA  
GCACTCCANCTTTGTGTTACAGGCGATAAATGGGAAAGGGGTAAAT  
>Contig48  
CATTCTTAATTAGAGAAACGCTCATTAAACTAGACACCCAAATTCTCTGG  
GGGGGATCATTCTTACAAGCATGCCCTTCTCTCTTAAAGAGAGAGCACT  
TTTTTCGCAAATAATGCTGCCATGAACATACGGGGTGCATGTATCTTCGT  
AATAGAATGATTTCTATTTTGGGGGGTATGTACCCAGCAATAGGATTGCT  
GGGTCAAATGGTATTTCTGGTTCTAGATCTTCGAGATCTTCCACACCGTC  
TTCCACAATGGTTGAACATAATTCACATTCTTACCAACAGTGTGAAAGCAT  
TCCTATTTCTCTGCAACCTCGCCAGCACCTGTTATTTCTTGACTTTTTAA  
TAATCGTCATTCTGACTAGCATGAGAGACAGTATCTCGTTGAGGATTTGA  
TGTGCATTTTGCTAATGATCAGTGATGTTGAGCTTTTTTTCATATGTTTT  
TTGGCTGCAAGAATGTCTTCTTTGAGAAGTGTCTGTTTCATGTCCTTTGC  
CCACTTTTTAATGGGGGTTTGTTTTTTCTTGTAATTTGTTTAAAGCTCCT  
TATAGACTCACAATAACAAAGACATGGGATCAACCTAAATGTCCATCAAT  
GATATAACGGATAAAGAAAATGTGGTACATATATACCATGGAATAGTATG  
CAGCCATAAAAAAGAAATGGGATCATATCCTTTGAAAGGACATGGATGAGC  
TGGAAACCATGATCCTCAGCAAACATGCAAGAACAGAAAACAATTGTTG  
CATGCTCTCACTTATAAGTGGGAGCTGAACACTGAGAACACAGGGACACA  
GAGAGGGGAACAACACACATTTGGGGCTGTGAGGGTGGGGGAG  
GGAGAGCATTAGGAAAAATAGCTAATGCATGCTGGGCTTAATACCTAGGT  
GATGGGTTGACAGGTGCAGCAAATCACTGTGGCACACATTTACCTATGTA  
ACAAACCTGCACATCCTGCACACGTACCCAGGACTTCAAATAAAGAGA  
GACAATACTTCTCCCTTAAGTGTCTACTGTTGCTTTGCAATAAAAAETTC  
CTGCCTTTCACCTCACTCTGACTTGTCCTGAATTCCTTCTCGTGATGGT  
GTCAAGAACGTGGACACTGGCTGGGGCTGGAGACTCACCAGCATCCGGAG  
ACCCTCCTGAGCCCTCCAGCAATACAACCTTTGACACAAACTATGAAATCA  
CAGATCCTCAAGAAAGCTCAAAGAACCACAGCAGGAAACATGATGAAACTA  
CATGAAGGAACATCAGAATTGAATTGTTCAAAATCAGTGATAAAGAGTAA  
ATCTTAAAGCAACCAGAACAAATATCCATCATATACGCAGAAATAAAG  
ATAAGTATGACAGCAGATTTACAAATAGAAAAAAAACAAGTGCAGCAAC  
AGAAACAAACTATCAATCCATAATTCTATACCTAGTGAAATTTCTTTCA  
AAACAAAGGTGAAATAAAAAAATTATTTTCAGGAATACAAAAGCGAAAAA  
ATTAATCACTAGCATTTCATCACTGCAAGAAATGTTAAAGGAAGTCCTTTA  
GGCAGAAAGAAAATGATACAAGGTGAATATTTGGATCCCTGCAAGGAACT  
AAAAAGATCCAGAACTGATACTTAATGGGTAAACATGTAATTTTCATCA  
ACAAGTGAATGAATAAACAAATCATGATATATCCATATGATAGACTACTA  
CTTAGAATAACAAAAGAAGAACTACTTATGCATGTGATAACATGAATGATA  
TTCAAAATTATTATTGAGTGAAAGACACCAGATCAAAACAAAGTACATAC  
TGTATGATTCTGTTTATATAAACTCTATAAATTGCATGCTCTTCTATAG  
TGACAGAAAGAAGATCAGTGGCTGCCTGCAGACAGGAAGAGATTACAAAC  
GGAAATGAGAATTCCTTAAGAGATGATGGACATGCTCATTACCCATCATA  
TGTATACAGCCATAATGGTTTTACAGATACATATATGTACACGCCAAC  
ATAAATATAAGTTATCAAATTACAGTAAGTTCTGACTTAATGTCACTAGG  
TTCCTGGAACTTTGACTTTAAGCAAAATGATGTACAGTGAAACCAATTT  
TACCATAGGCTAATTGATATAAAGATGAGTTAGGTTTTTGGTTTTTTTTT  
TTTTGACATGAAGTCTCGCTCTATCGCCAGGCAGGAGAAGAAGAGTTAG  
GTTTTACAGCATGTTTCTGGTCACAAGAACATCATCAAACCTGTAAATAA  
AGGCACAAAACACTTCTAATATTAATATCAAATAAATATGAGTTATAC  
AGAATTTAAGAAAGATTAATAAAAAACAAGTAAATCATTATTTATGGGAT  
TTTTGGTAATCAGTGAGTTATGTGGTCATAGTGAAGTGGGTTAAGTCAA  
GAAATAAATGTTTGCAAAACAAAATTTTAAAGATCCTCTCCTACCACCA  
CACAAAAACAAGAAAACACGGTGGGCTCGCTAAGCACTTTTGTACCCT  
CGTATCTTATGCGTTTGTATGATTATTGTAAATGCTTTATGATAATTTTT  
AGAGACAGGGTCTCACTCTGTGTCTCAGGCTGGAGTGAAGTGGTGCAATC  
ATAGCTCACTGCAGTCTCAACCTCCCGGATTCAAGAGATCCTCCCACCTC  
AGCCTCCAGTGCTAGGACTACAGTTGTGTGCCACCATGCCCATCTAT  
CTTCTTTTTTATTTTTTGTAGAGACAGGGGTTGTGCTTTGTTGCCAGGC  
TAGTCTTCAACTCCTGGGCTCAAGCAATCCTCCTGCCTCAGCCTCCCAA  
ATGCTGGGATTTCCGACATGAGCCAGCAGCACCTTGCCAGCATTTTATT

TCATAATAATTATAAGTCATTTCCTTCATTCATCTTACAACCCACTTGTTTC  
CAGTTTCAGGATCTCGGGTGACCAGAACCTATTAACGTTACGCACAAGTC  
AGAAACCAGCCCCTGGACAGGACACCATCCTACCGCAGGGAGAATTACAC  
ACCCCACTCACTCAGACTGGGACCATGCAAAGAACCTAACGTGCACTTT  
GGAATGTGTGTTCCATACCCACTAGAACAGCTAAAATTTAAAAGACTGAC  
CATACTTGAGTGTTGAACAGGATGTGACACAATAATCTTTTAAGCGCT  
TCGCGTAAATGGCACAGCCGCTTTGGAAAACAGTTGGCAGTTTTTCAAG  
TTAAATATACCCAACTCTATGATCCACTTCTCAACAATCAAACAAGAGA  
AATAAAAGCAATGTCTACACAAAGATGTATACACAAATGTTTCATTGCAGC  
CTTAATTATACTAGCCCCAAGTTGAAACAAGCCAAATGTCCATTACCAGA  
TGACTGGAACATACAAATTGTGGTATATTGATACAATGAAATACTACTTA  
GTAATAAAAAAGAAAGAGCTATTAACATAAGCAACAACATGGATGAATCT  
GAAAACAATTATGCTAAGTGAAAACAGCCACACAAAAGTTACATACTGTA  
TGATCACATCTACATAAAATTACAGAAAAGGCAAATAATCTATAGACAG  
AAAAGCAGATGAGTGGTTACCTAGGGATGGGGCAGAAGGGACGAAAGGAT  
GGATTGCAAAATAGCACAAAAATATTGGAGGGATGACAAATATATTTCATT  
ATCTTGATTGTGGGGATAGTTTAATGGGTATATATAGAGATCAAAGCTCA  
TCTAATTATACACTTTAAATATATGTATTTTCATTGTGCATCAGTTATTCA  
TCAACAAGACTATAAAATAATATATGCCTACATACATTTTTAAATATTCA  
AAATCTCACAGTTATATACATAAATGCAACTGAATATGTATTTCAGATGTT  
TTAACAAGCAGAAAGGACTGATTAACTCATGACAGCGGCTGTTTCTGGG  
AAGGGTGTAGGAGACAAGAGATGGAAAAGAGGATGAGAGCCAGAAGAGAC  
CCTTGTAATGTTTCCTTTCTTTTAGTAAAAATATATTGACAGTTAAAGCT  
GAGAGGTGAGAATAATAGTCTCATGGCTTTTGTGTCCTTAAAATTTTACA  
AACTAAGTGAAATGGGAGAAAGCAAAAAATAAACTTAAATAAATGTTAT  
ATTGCCCAAAAAGAGATTTAAATGGAGGTTAGACACATGAGACTTACGT  
TCTCAAAAAAGTAGAATCTGCAGGGAAGTTTAACTATAAAGAATTAA  
AATCTAGCTTCTACCAGCCCAAGCCTAAAATGTTCTGCTTTATTCTTCC  
TTATTATAATTCATAGGTAATATATTTTATGTTTGCAAATGAATGCAGTG  
ATATTAGATCTCTAAGAGGTGCTAAAAATGAAAAGTACATATTCCAATTT  
TTCCCAATTTTCTTCTTTCCATGAATGAAAATATACATATTTGATG  
ATTTCCAAGTTTATACAACCGATCTTTCTCTTAGTTTCTCTTACCAAAT  
TCCCTCCCTCACTCAGCCACCAGCCAGTCCAATGTGCTACCTGCACAGC  
AGCCCTCATACCATCCACACTCTCATCAGGATCCTGCCTGACCTGCGAGG  
AGCAGCAGCAAGAAGGAGACAGAACCTCCACGCTGAGCATCTCAGGGCTT  
TCTCAGAGACTCCAGAGGACCCTGATAGGGACAGAGCCTGGCCAGCAATC  
CATGCTGCCAGCTGTATGATTGTGGGCATGTAAATCTCAACTGAAAATG  
GGTGTAAATAACATGTTCTTCCAGAATGAGCTTTATGAAGATCATAT  
AGCTGTTTGGAACTCAGACAAGCACTGGTAGGAATACAAACAGGGGAGCC  
AACAGCCTATAAATAATACTTTAAGAAAGGGCATGAATGTAATTACTTAG  
GAACAAGGCAAGAGTGGAGAGATGCCTAGGACTGAGCTGGACAAGCTGC  
ACCTTTTAGTGGCTCAGCCCATGGGCTGACAAGGAAAATGGAGGAGCTAC  
CAAAGAAGGTGGAAGGATTCTGGGAGAGTGGCCCTCACCTGCCCAGGGC  
AGGGCTCAGTGGGAGAGAGGGAGATCTGTTATAAATGCTGCCAGGAGGTC  
GAGTCATGTGAGAATGTCCATGTGAAAACATCCACTGTGTGTATCTAAAG  
AGAGTGGCTGTAAACAGGTCAGGGTCAAAGGTCTTATTGTCTCAGATGT  
TATCTGCATGCATTGTCTCACGACCAAGAAAATAAGGAGCATGGACACA  
AAGGGTTAGGTTGAAGCAAAAATTTAATAAGTGAAAGAAGAAGGCTCTCT  
GCAGTGGAGAGGGGAGTCTGAGTGGGTTGCCACTTTGACAGCTGAATCCA  
AAAGCTTTTATAAGAAAATCTTCTCATATCTGCAGCTGTTTGAGTAATTT  
CTCTTACCTATAAACTGTCTGTATAACTCTCCCTTATCTATGCAGCTGT  
GGGATGTCTCCAGGTAAGCATAAAGTGTAGCTTCTCTTGTGTTGTATAACT  
GTGGGTTTGTGTTTAGGCAAGCCCCATCCCCTCCCTGTGTAAGCTCCCAT  
GGAGCCCACCATGTGCATATCTGAGAAGTGGAGGAAGCTTTCTCTGGGAG  
CTCACTGATCGTACAAAGAACAAGAGGCTTCTGTGCCGCTTATCTATTCA  
GGTGCAGCCTGAGTTTTCCCCAGGCTGCTCTATTTTTGCCTGTAGCTATG  
ATTTTTCAGGCAGGCTGCTTCTCTGAAGACTAGCCTTAACTGTCTACCTA  
TCAGATTTTTCTTTTCTTCTCCCTCAGCTGGTTCCCCTCACCAAGGCTG  
AGCAAGTGAAAAGGAGGGCACAGGGCAGGCCAGTAGTGAGCAGCAACAAG  
GAATAAGACAGCAGAAACCACTCTTCACACCTGGGTTGAAAGGGGTGGG

GAGCCAGGACTACAGC<sup>1</sup> LAGGTAAGAACATAGGTAAAGAGATACTGTTGT  
TGTGTTGTTTTTAACTATGAGAAGCATTGAGCTTTAAATTTCTACAGGAA  
GGATCCAGTTTACAGACAGGAGCACCCAATATTCAGAAGAGAAGAACATGGT  
GTAAAGGTCCTGGGAAGGCTGAGAGGATTGGGACTCAGAATCCAGAGCAG  
AAGCCGTCTGTGAACAGAAGAAGGACCTCCCCAGTGTAGCAAGAGGGAG  
GGAGGAGGGACAGATGCCAAGATGGTTCAGGAAGAAGGTTTGGTGGTAAA  
TGTGAGGCTGTGCTCACCTGCTGGCTTCAATTTCTCTTTAAATGTGAG  
ATGGAATCATTTGATGAAGGCCATGCCATGCAATGAAATGGCAGTCTGAG  
GCATGGAGCAGCTCCAGCTTAGCCCGTGTTTAGGGTAATTATGGCTCCAA  
CCCAGGAGATGAATATGACTAGGGAAAGTGAAGTCCAAAAACAAATGGTC  
TCAAGTTGACTGTGAGTCTTCTGGGAGGCTGAGACGACAGGTGGGGTTGA  
CAAGGGAAGGGGAACCCACCTGCTGAAAAACATCAGGCTGTTGGCTGGGG  
GAGGGGTGAGGCCCTGTGTTGTAGAGATGGATGGATGCCTAAAGTTGGGTA  
AAGGTTTCAACTCTACCCTCTGCTGGGTGTGGAAATAAACAAAGACCACC  
CAAATGAGAACAAACAAAGACTATTTATCCAGAGCTTGCTCTGACAAGGG  
AGTCGGCAACCATCACTTGCTTGGCAGAGACTCAGAAGTAAGCAGGGGAG  
AAAGCCTCATAGCAGAAAGAAGGGAAAGTCTTCATGTATGCCCTGAGTGGC  
AGCTGTAGATGTGGGTGAGTTGCAGGTGGCTAACTAGAAATGGGGGACTC  
CTGTGTGATTGATTAGGAGCATGTTTGGCTTTCTCTGGTTGGTCTTACAT  
TGGAAAGAGGGGAACAAAAATTTAGGGCAGTTGTGAGTTATTAATCAAGTG  
TTGGC<sup>2</sup>CAATTTTGTGACTGACTGTTTACAGGAGTGACTGGCTCCCTGGATTGT  
TTGCTAGAAATAGTGGTCTTCACTTCTGCAAGTCTGACTTTCTGGTAAT  
AGGCTTCTGGGTGGCTATTGTGGATAATAAGTGGGTTTCTGAGCTGA  
TTTCTGCAGATTGTGGATCAGAGTTATTTTATATAACAGTCTGACCATT  
TTCCACTGGCATATTCCATCTTCCAAGAGCTGGCCAAGCTGCTGTCTTAT  
CTGTCTCCCCCAGCCCCCTCCACTCTGGCTGTGAAAATACAAGCCACTAGG  
TGAGGAATGGGGACAATTGAAGACTGAAAGCTTTTCTTTGCTGGGTTTCGC  
AGAGCTGAGGAAGAAATGACAACATCCAAGTGTCTGCCCTGGGCCAGTT  
TTAGGACTGTAGTGGTAATGCAAGGACTGTGTGAGTTTATATTTTCATTT  
GTCTCTCTAAGTAAAGGTGGAAGAAAAAACAAGAAATTTGTCTGTCTGCA  
GTCTCTGCAAAAGTCTAACACTGTGCTTCCCAACATTGCAGCCATTAGCC  
ACAGGTGAGTATCAAGCACTTTAAATGAGACTGGTCCAACTGAGATGTG  
CTCTGAGAATAAAACACACAGCAGATTTCAAAGACCTAGTACATGCCCTG  
ATTTCAAGCTATATTACAAAGCTGTGGTAATCAAAACAGTATGGCATTGG  
GAAAAAATAGACACATTGGTCAATGTGACAGAATAGAGAGCCCAGAAAT  
AAACCCGTGCATGTATAGTCAACTAATCTTTGACAAGAGTACCAAGAATA  
CACAATGGGGAAAGTCTCTTCAATAAGTGGTGTGGGAAACTAGATATC  
CACATGCAAAAGAAAGAAATTAGACCCTTGATTACACAAAATCTAAAT  
TAATTCAAAATAGAAAAAGACTTACATGTAAGATCTAAAACCATAAAACT  
CCTAGAAGAAAAACATAGGGAAAGAGCTCCTTGACACTGGCATTAGCAGTA  
ATTTTTTCAGATATAACATCAAAAGTACAGGCAATGAAAGCAAAAACAAGT  
GAGAGTATATCAAACTAAAAAGTTTCTGCACAGCATAAACAATCAACAGA  
GTAAAGACATGACGTATGGAATGAGAGAAAATATTGACATCTGACAAAGG  
GTAAATATCCAAAATATATAAGTAATTCACACAACCTCAGTAACAAAAGCC  
AAATAACCTGACTTTTTTTTTAAATGGGCAAAGTACCTGAATAGGTATTC  
CTCAAAAGAAGACATACAAATGGCCAAGAGATGTATGAAAAGCTGCTTAA  
CATAACTAATCATCAGAGAAATACACAAATCAAAACAAGATATCATCTCA  
CACCTGTTAGAATGGCTATTATTAAAAAATGAGATAAGTGTGGCCAGGT  
GTGGAGGAAAGGAAACCTTGTACATTATTTCATAGGAATGTAAATTAGTA  
CAGCCATTATGGAGAACAGTATGGAGATTCCCTAACAAAATTAATAATAG  
AATTACCATATGACCCAGCAATTCCTTCAAGGAATACATTCAAATACT  
ATCAGTATCTCAATAAGATACTTGCACTCCTATGTTTCGTTGCAGCGTTAT  
TCACCATAGCCAAGATACAGAAACAAGTTAAATGTCCATCAACAGATAAA  
TGGATAAAGAAAATCAGGTACATATATATATAAATGGAATATTATTTCAG  
CAAAATCCTGACATCTGAGATAACCTGGATAAACCTGGAGGACATTATGC  
TAAGTAAAAATCAAAGCCTGACACAGAAAGACAAATACCACATAATCTCAC  
TTACATATGAAATATGAAAATGTTAATTTTATGGAAACAGAGTAGAATGG  
TAGTTGCCAGAGCCCTGAGAGTAGAGAAAATGAGATGCTTGTCAAATCAAA  
TCATCACATTGAATATATATAATCTATTTGTCAATTAAATATTTTAAAGAA  
TAAAAAATACCTGGCACCAAAAAAAGAATGCAAAATGTCTCAACAATGTT

ATATGTATTGCATTTTG..AGTGATAATAATTTGAATATTAGGTTAAATAA  
AATATATTTGAAAAATTAACCTTACCTATTTCTTTCCATTTTGTTAACA  
TAGGTACAAAAAAATTAATAATTACCTATGTGGCTCATGTAGGTGGCTC  
ACATTATACTTTGATGACACTATACAGGCTGGTGACCATATATCTCTTAG  
ACTAGTCTAAGTGATTTAACAGTGGTTCCAGAAAGATCCAGGTTTAACAC  
CAATGAAAGGGCCAGCTGGCTTAGCCAGCTTGTGTGGGAAATGTTGGGG  
AGTGGTTTAAAGACAGGGAAAAGCAAACTTTTGATGCTATTGACTTTTTG  
AAAAATCTTTTGTGGCTGAAAAACCAAAACATTATT

>Contig49

GCTCGAGTGTGTCTCTAAAGCCTTTCCCCCATTTGGCTCCACTATACGCAC  
TCTCCTGGTTTCTCCTCCCTCTAGCCGCTGTCTTTGGTCTCCTTTCTGATT  
TTGCTGCGTCTCTGTCCCTGAATGATTGCTTCTCCACTACGGGGTGAT  
TTTGCTCCCCAGGGGACATTTGGCAATATCTGGAGAGGTCTATGGTTGTG  
TTTGAGGGTGTGTGCTACTGCCATCTAGTGGGGAGAGGCTAAAGATGCTGT  
TAATGCCCAGGACAGTCCCCATAACACAGAATTATTGAGCTCAAAATATC  
CATGGTGCCAAGATCAAGAAACCTGCTCAAATATTAGCATGTGCTGAAG  
GCCCTTCTCTTTCTTTAGCAATATCTGCCTCCTTAGGGATCTTTTCTAG  
TCTCAGTGGTTTAAACATTTAAAATCCCAAATTAGGCAATAAATTGGGCCC  
CAAACCTCGTTAGTATAAAATGTAGAACTGTGTTATTAGAAGGCTAATAA  
AATGACCTGGTGAGCATCTGCAGCTAGCCTCTGAGCAATTCTGGGGACCA  
CGTGCAAGATAAAATCCATCTGTTCCCTCTCTGTAATGTGGCGCTACCTTG  
TGGCCGATTTTCTCCTCGGGTTAAATATCTCTGGGGATGCAACTTGTCTG  
GTTAATGGCTGTGTGAGGCCAGCGCTGGTGATAAAGGAATCAATCAAGA  
CAATATTGAATTTAGAAAGGCAGATTTATTTAGAGAAAAGGAGAGATACG  
TTGCAAGGGAGCAATGGGCAATACAGCAGAGGGAAGGCTGTCTGCAAAGA  
GGCAAGGGCTACGTATGACGTAGGGCTGCTTAGGCTGAATGCTTGACAGAC  
AAGATGCTTGGCTGCAGGTGGGCTGTGAGCTGAGTCTTGGGTGCTAGTG  
AGCCATTGGCAGCTGACCTTATTTCTTGAACATTCTGCTCCCTGCAAGCA  
TTTTAATGTTAAACCGCCAGGTGAGTTTGAATTTTCTTTTCTTTTTTTT  
TTTTTTTTTTTTTGCCTTTTAGTAGGACCTGCCGTTGTGAGACTATCTGAGG  
TAAATTAGACACCCTCCTGGTTTAAAGTCACCGCTCCAGTGACTAGGCAGG  
GAGCTCTTCTTGAAGAGGGTGTGGGCAGTGGGTACTTTGCATGTTGTCC  
ACACCAGGCGAGCTGCTGCTTACGGGCTTTGCATTTGCTCTTTTCTTTG  
CCCAAAATGCACTTCTCTCACTGTTTACATGATTTTCTCCCTCTTTTCC  
TTTTAGTCTTTGCTTAAATATCACCTTCTAGGGAGGCCTTCCCACACCAC  
CTCTTCAAGATTTGAGGGTATGCACCCCCACCCCTAGCCTTCTTATCCCT  
CTCCACTGCTTCTTCTCAAAGCACTTGTTACGTTCAAATAAAATAGATT  
AGTTACTTTTACTAGTTCTAATTTTACTATTTTGTGTTACTTCATCAATAC  
CCATGTAATCTCTGGAAGGAACGTTTCTTTTTGTAGTGATTTCTAGCAC  
CTAGAACAGTACTTGGCAGATGGCAGGTGTTCAAAGTATTGTTGATTA  
TTTTCTCAAAGGGCATGGAGTCTTAGAAGTTGAGAACACAGTTCTAAGC  
ACAGCTGTTTAGAGACTATGGATGATGCTAATGGCTGTATTCCCAGTAGG  
TGGGGCAATTCTCAAATTGACCTGGAATCCTTGAGATCTGGGGACAGTCA  
CCAAGCACTGGGCTCTGTGGGGAGAGATGTGCTGGTTTTAGAGAGGAGA  
ATAGCATCTGGGGGACTTGGCCCCAGGGCTTTCTGTCCCAATCTCTTC  
CCAAGTACTGCTCCAGAGGCAGGAGGCCTTGTCTGTAGCTGGTCAGTCTG  
TAAGTGTTCCTTCCATCTACACAGATGCAAAGAAGGCTGAGAAAAGCA  
AGCTGTGAGGTGAGCAGGGCCCTGACTCCTCCCCAGAAGGCACTCAGAA  
CTTCCATAGGGCAACTGGAAGAAGGTTTCTACTTCTCACCGGCAGCTGT  
TGCTGGGGAAAAAACAGCCTCAGGCCCTACCCTGTGCTGAGAACCTGAA  
TCCAGTATCAGGTTCTCCAACAACTTGGATCCAGCTGACCCTCACAAGG  
GGTCAGATGCAACCTTGTAGCATATGGAAAATGGCAGCAAGGTCCTTGTG  
TGGACTATGCCTAGAATCTAAATTAAGACAAGGCCTCAGAGGGGCTAAGT  
GACATCTGTCTCAAAGTTTACAGCTAGTGTGTGACTAAATCTTGATT  
CACCTCTCAGGTTTTACCATATCCCAAAAAGGTTGAAACAAGAAAAG  
TTATCTTTGGGCAATTACCTCTTCTGTTCTTCTTACCTACTAATGT  
TCTAGGCTCACCCTCTGGTCTGCAATCTCACTGAACTGACAGATCCCTCA  
TGGCCTAAAGGGTTTTTCACTGGGTTGACTAGGCTCTCCCATTTGCCCTGT  
CCTACTGTCTAAGGCACCTCCTGGGTAGGGTGCCAGCGTCATTCTGATG  
CTGCCTGACTTCTTCCAGCTACTTTTGAACTTGGTATCCATGGCAGA

GGCTTAAAGGGCATGTTCAAGGTACTTTTATTTCCAAATCCCCAGTGGC  
ATCAAGGAAATCAGCATCTCTGGATAGCTCTACTAAGGCTTAGTTCTCAT  
TGTCCAATCTAGCTCCTGGGTTCATGGGAGGCATTGAGGAAATATTTGAGT  
GTAAGAGTGAGTTGCTTTACCTCCAGAATATCCTTCCAATGGCTCTGAAG  
CAGGCTGTGGAGTCTGCTGGCTGATCACAGTTCACAGGTGGCTCCCAA  
CCTGTGGTCTACATCCATCCTTTGTGAGTGTCACTGCCATTGTCCACAA  
ATGTCATTTGGGCCTAGCCCCCTGGGATAGTAATCAGTCTTTACATAGATA  
TACATTGTGCTTTACATCCACAGTAATTCTGAGTGGACCTTAAATAAAT  
TCCATGTGAGGTCTCACCAGCCCATGGGTTACAGATGGGGTTACCTTTCA  
GCCTTGTAAGGTGCCCCGTCTTTGAGTGTAGACATGGACTCACACAGAGT  
CCACTCCTGCTGTTCTCTGCTCTTGCTGAGGCTTCTGCTGCTGCTGCTG  
CTGCTTTGCAGAGGCTGGCCAGCTGTGGTGCCTGAGGCACCTGTGTCTTC  
ACAGCACCAACTTGCATGGTGGCCACGGTGTAGTTGGAAAGGGATGCTTA  
GATGGGAGGCCAATGGGAGCTGCTTCAGGAGGCAAATCCAAGTCACAGAG  
ATCGAGTCACCGAGAGCATAGTAACTCAAAATCCCTTCTTCTGCTTAAT  
AACTGAGATGCTGTCACTGGGTTAACCTCACCAGCCTTGTCTGCTCTTC  
ACTTAGAGTGATTTCTGTCTTAGAAGGCTCCTCATATCCTTCTGGGGAAG  
GCTTCTAGTGTGATCCACAGATAGCTGGACCAGGCATGTCCAGAAATAATC  
TGATTCTCACATTTGAGTTAGCCAGCGTTCCAGCTATATCCCCATTTTG  
TGTCTATATAAGTTACCAAAGCCCCACAAGGATATTAGGTGGCTCCTTAGT  
TTGCTTTATGATTATGCCTTGTGTGTGTGTGTGTGTGAGTGTGTACGCCT  
ATGAGGATTCCTTCTCTCCCGTTCTTGCTATGGCTTCTCTTCCCCACTGA  
TGGGCTGTAGTTCCCTGTCCTTTTGACTTTGGGCTTAGTCATGTGACTTT  
TTTGCCAAGGGAATGTGGGCAGAAAGTAAGTGGGAGCCAGTCCCAAGCTAA  
GGCCTTGGGAAGCATGGTGAGCCTATGCCAGCTCCCTCAGAACTCCTTCC  
CTTGCCCATGAAGAGAGAATAACCTGGATTGTACCTTCAGCCCCATGTCCT  
AGAATACAAACATGGAGAATAATGAACCTGACTCAAAGGCTGAAGGGCAG  
CTGAGCCCAATGAGGTCAATTGAAGTGCAGCTACCTACAGACCTGAAAG  
TGAAATAAACATGTATAAGTCTCTGACGTTTGGGGTTTGTGTACATAGCA  
TTATTGTAGCAGAACTTAAATAATACTGGGGGCTAAATATAGTGGACCA  
GTGACAGCACAGAATGGTAAATGGAGTGATTGTTACTTACATCACAACC  
CTTCATCTCTGTTGATGGACACTAAAATCAAAGTGGCAATTACTCAGAGT  
TGGGAGTCATTGAGTTGCATCATTGTTGTTTAGAATCATTGACAGTTTGA  
GCTCTAAGTGATTACAGAGATGGTTTCTCAGCTACAGGTAAATAAACAA  
AGGCACAGAGAAGTAAAGTGACTTCTAGAGGGCTTCATTGATATTTAGCA  
GCAGAATCAGAGCTAAACAATGAGTCTCTCATCTCCAGCCTTCTATTCT  
TGTTTCTAGGTTGGGATTTTGGGAAATAGTGCAGAGAGATTAGCAGTAG  
TGACATGGAACAATGTGAGCCTCAGCTTCCATCCCTGAGGCTGCCTTCAT  
CTGCCAGGGAATGTCTCTGTGTGCAGCCTTGCCCTCTGCACACAGTGTG  
TATGGCCACCTGAATAAGTGTCTTTTATAGCGACTAATGGATTGAAATG  
GGTGCTAGAGCAGTGCTTCTAAAACTCCATGTATTAATCATCTAGGGGT  
CTTACCAAAAACGCATGCAGATTCTGATTGAGTAGGTCTGGAGTGGGGCT  
TGACATTCTGCACTTGTAACACATGGACCACACTTTGAGTAGCAATGTAT  
TAGATCATTCCAGTGGAACATGTATGAGTGATGGAATGAACAGATATAA  
TTAATCCAGGTCTGGTAAGTGAGGTACTGATACATATTAAGTTGAAGTGA  
ATTTACATCAAAAATAATGGTTACACAGTGACTTTTACTGCCCCCAAT  
TCTTTCCTTTTGTAGTGGTTTCAAAGTGAAGTGAAGCCAGCCAGGTAAAGTC  
CCTGGTTTGTAGTGTGATTAGAAGATTGATCCAGCTTTCTCTCTCTCT  
AATTCTTTAAATATGCAATGGCCTTCTAGAACTTGTCTCTCAGGCTCCC  
CATGAGCCACCTGTCTTAATATCTTCCCCCAGGACATTCTCTGGGTCA  
AGGAAGGAATCAGGGACTAGGAAAAGTAGAAAGGTTGCCTGACAGTGAGA  
AACTTTTTGCACTCCTATTTGTTCAATTCTAAAATGTGGGTATTGTTGGG  
GCTTCTAATTGGAATCTAACCTGAAATTCAGGCATGTCTAGCTATATATG  
ACCAAGAATTAGGATGAGTTCCTAGAAAGCCTATTTTTCAGGAGAGCGGTC  
AGTTAAATTGAAGTTTATGGGTTTATGGTAATGGGTTGGGGAGTTTACTT  
CATTAGCAATAGCAACGTTTTTGAATCAGAGAAGTGATTGTAACACACT  
GTACATAGTTTTCTCACTTAGATTTATCTCTGGGTCAACCCTTGTGTGGAC  
CTATATTAGAATCATTTAGTGAAGAAAAGGTGGGTGTCTTAGGAAAAGA  
GCCATTTATTCAAATGTTCTGTTTGACATTAGGGCACTGGCAAGACTACA  
GAATCAATAGATATTTAAAAACAGCCAGGTGCGGTGGCTCACGCCTGTAA

TCCCAGCGTGATTTGGGCTACTTTGGGAGGCTGAAGCGGGTGGATTG  
TGAGCTCAGGAATTCAGACCAGCCTGGTCAACACGGTGAAACCCCTATCT  
CTACTAAAATACAAAAAATTAGCCGGGCATGGTGGCAGGCGCCTATAATC  
CCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCG  
GATGTTGTTCATGAGCTGAGATCGCGCCATTGCACTCAAGCCAGGGCAAGA  
ATAACAAGACTCTGTCTCACAACAAACAAGCGAACATACGAAACAAACGT  
AACATCCAAACTAGCAGGTACATGCCGTGCCAGTCATGACCCATGGTCAT  
AAGATGTCTACAGCTCAGGAAGCAGCTGCACAATGCCTGCATAGACAAAC  
TCTTATGAAAGCAGAATGTCCTGATGTCTCCATAACACATAACAGTGTAT  
GCTTTTATTATGGTCATACTCTAGCTGTGATGTACCTACGCTCTAATATG  
CCAACGATAGTTTTCTTTAAATCATCAACATAATAAATGTCATGCTGTCA  
GTCCCCCACATGTAGACATAACTTAGCTGGTACATGGATAAGAAACCTAT  
ATTAGATAACCTTAGGCCAGGTGTGGTGGCTCATGCCTGTAATCCCAGCA  
CTTTGGGGAGGCCGAAGCGGGTGGATCACGAGGTCAGGAGATCGAGACCA  
CCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAA  
TTAACCGGGCATGGTGGCAGGCACCTGTGGTCCCAGCTACTCAGGAAGCT  
GAGGCGGGAGAAATGGCGTGAACCCAGGAGGCGGAGGTTGCAGTAAGCCGA  
GATCACACCACTGCACTCCAGCCTGGGGGACAGAGCGCAAGATTTTCGTCT  
CCCAACCCAAAAANCNANNNNAAATTTGCACCCAAATCTGACTAATTCCA  
GAGCCAATTCGAATTTAGAATCGTTATATCTCCCTGGTGAAGTGAAGCTT  
TTATCTTTAAGGAGACACACTCTTTATGTCTACCAATGCTTATTGECTTA  
AAGTCCACTTTGTTCAGATACAGCTGCTTTCTTTAATTAGTTTTTGTGTG  
GTATATCTCTTCCATCCTTTTTCTTTTCAGCCTTCTCCATTCTTACATTT  
TAGATATATTTCTTTTTCTTTTTTTTTTGGAGAGAGTCTCACTCTCTC  
GCCCAGGCTGGAGTAGTGCAATGGCGCGATCTTAGCTCACTGCAACCTCC  
ACCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGG  
ATTACAGGAGCCCCACCACCAAGCCCAGCTAATTTGTTGTATTTTTAGAAG  
AGATGAGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCA  
GGTCATCCACCACCTCGGCTTTCCCAAAGTGTGGTATTACAGGCGCGA  
GCCACCATGCCAGCTGATTTTAGCTGTATCTCAAAAACAGCATGGGTTC  
TGTTTGCTTTCCTTATTTCAGCTTTATAATGTAAATCATTTACATCAAACA  
TCTAATACACCATGGACTGTAAAACACAGCCATATTTTATGTATGAATTA  
AAAAAAAACACCACCAATTAGTTCCTGAGACACACACCTTAACAATAT  
TCTCTGTAGTGTGCATAAATCAATCACATCAGTTTCTCTGCACCTCAAAT  
TTCTTTCTCAATTTCTCAGAGATATGGCAATTTCTCTGGTTTTACATTCC  
CAGAAGCAAAGAAAAAGTACACAGCTTCTTCAAGTCATGAGTAGCTTCTT  
TTTTATAGCTCTTGGTGTGTTGCAAAAAGATTGGAATTGCTTCACTAATA  
CTAAATTTTCATTCTGCTGCTCTGTTTCTATGACAAGTCAGAGGGCATCT  
TTTTGAAGACATTCTAAACAGCAATTAACTCAAAACATGTAATGACAAT  
GACACACAAAACCTCAACTGATGACCAAATGAAGAGTTCCAGCCAAGTTGA  
CACAAGCTGGCTGACAGAGCTTGTAATACACACAGCTTGGCATATGCCTC  
GCCATTTAGAGATGTAAAATAGGAATAAATGTTTTCCCTTAAATCAAT  
GAAATAGAGCATTGTGACTGAAAATCTACGACAGTTATAGTGTTTTCTAT  
TCATTATTCTCATTCTGTTTTCTTCTCCCCCTTGCTTTCTTTAGTTTGAA  
TATTTTCTATCATTTCATTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
TATTTTCTATTTTTTAGCACTTACCTAAAATTACTCTGTAATCCATGGAT  
CCTTAATTTATTTAAAAAACTAATGTAAATGAGTAGCTTTATTTTCTCTCC  
CATCTAATTTAAGGCCACAGAACACCTTCACTTACCTCAATCCTCTCCC  
AACTTACATGCTTTTAAATGTATATATGTTAATACCGTATACTTTTAAAA  
CTTTCTAAAATAGCATTATTTTATAGCATGAGTGTTCAATTACATTTTTG  
CATATATTTAGAATTTTCTTTGCTCTTCGTTTCTTCTTCTATTTATGACT  
CCCCTCTGGGATCATTTTCTTCTACTTGAAGTACATAGTTTAGAACTGC  
ACTATTCAATACAGTAGCCACTAGCCATGTGTAGCTATTGAAGTTTAAAC  
TAAGTAAAAATTGAGTAATATTTAAAACTCAGTTCCTTCACTCTCACTAGCC  
ACATTTCAAGTGCTCAGCAGCCACATGTGACTAATGACTACTGTACAGCA  
AACATATAGAACATTTCCATCATGGCAAAGAGCTCTATTGATAGTGTTCA  
TCCAGAGTTTCTGTTCAGGACCAAACCTGAGGGTTGGGCTGCTATTTCTC  
ATGGCCCAATAACAAGATGCAGATGAGCTGGGGAGGAAGAGAGTTTTTAT  
TTCTGCAACCAGTTACAGGGAGAAGGCCTGGAAATCATCACCAGGCCAAC  
TCAAATATGACGTTTTTCCAGAGCTTATATACCTTCTAAGCTATATGTC

TACGTGTAAGTGTGCATLACCTGAAGACGTAAGTGATTAACTTCTTTTA  
ATCTGTAACCTAAGGTCTGAGTCCGGAAGATCTTCCCCTGGAGCCTCAGTA  
AATTTACTTAATCTAAATGGGTCCAGGTGCTGGGGTAATTACCCTTATCT  
TGTCCCCTGCTAAATCATGGAGGTTTGGGGAATTCCTTTAGAGCACCAT  
TAACCTGTTTGTGAAGGCCTGGGAATTTCTCCAAACCCCATTAACCC  
TGTTTAATCCCAAATTGGTTCCGTTAAAAATTCCCTCCTTAATTTGTCCA  
ATTTTAAAGGCCCAAAAAGGCTGGGGCAAACCTCCTGAATGGCCTTTGTT  
ACATTCCAACCTTTGTTTAAAAACACCGGTTTTTAATATTTAACTTAACC  
ATTTAATCTCTACTGAAACACTTGTATATAAATCTGCATTAATGAGAAC  
TGGCCTGCGCCATATCTCCTTCTCAGAATATCTTAGGGTTGTGATCCCCT  
GTGTGAAGAGAATATATCTCTGGAGATCTCAATCTCTCTACCCCAAAAAA  
AATCTCACTCGGAGAAACTCAGACTCTTATCTCCACAGCGCTATCTCTC  
TCCTCTCC

>Contig50

GCTTGTCTAAGATGGTGCTCCTTGTGCTGTGCCTGCTTTCATCCTGGGA  
TCTCCCTTCACCATCAGGATTGCCTTCACCTCATTCCAGTCTTGGATCTT  
TCTTCTTGTCTTCTGAGTATTTTTTTTTTTTTTTTGTCTGCATTCCTTCA  
GTGGCCTCTTGGGAAAAGATGTGTAGGGAGAAAAATTTCTTTAGAACT  
TGCATATCTGACAATATATTTATCCTATCCTGACATTTGGTAGATAGTTC  
AGCTGGGTACAGAATTCTAATTAATTTTCTTCTGATTATAAGACATT  
GCTCCATTTTCTTCTGGCTTCCAATATTGCTGCTGAGAAGTCTGACACCA  
TTCAAATGCCTGATTTTTTCCATGTGATTGTTGTTTTCTGTCTGGAGTGT  
TGTAGGATTGCCTCTTTATCTACAGTGTCTGAAATTCATGACGTAGGT  
CTTTCTTCATTCAATTATGGTAGACACTCAGTGGGCCATTTAATCGGGAAA  
AACATGTGTCTTCAAGTCTACAACTTTATTACTTCCTTTTTCTTGTTG  
TCTTCTCTGGTCTGTTTTTCAGCCCCGAGTCTCTTAGATCTGTCTCTAA  
TATTCCTATTGACTTTACTTCATTTTCTAAGTCTTTATCCTTTTGCTTTA  
CTTTCCGAGAGACCTGCTTAACCTTATCTCCCAACTCTTTTATTGAATTT  
CATTTCTTTTACTATATTTTTTACTTTGAATACACCTCTCTCTTCTC  
ACATTTTCCCCCATAGTATTTTGTCTTCAATTGACAGTCTACTATCTTA  
TTACTCTGGAGATATTAATAAGTTTTTTAAATTTTTTATTATTTTTATT  
TTCAAAACAGTGTCTTACTCTGTCACTCAGGCTGGAGTGCAGTGGTGTGA  
TCATGGATCACTGCAGCCTTGATCTCTGAGCTCAAGCTATCCTCCTGCTT  
CAGCCTCCCAAGTAGCTGGAACCACAGGCATGTGTCAACCATACCCAGCTA  
ATTTTTTTGTTTTTGAGGTGGAGTCTCACTCTGTAGCCCGGTCTGGAGTG  
CAGTGGTGCAATCTGGGCTCACAGCAACCTCTGCCTCCTGGGTCTGGTT  
CAAGCAATTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGAAACA  
CACTACCATGCCAGCTAATTTTTGTATTTTTGTAGAGACAGGTTTCACC  
ATGTTGGGCAGCCTGGGTCTGAACCTCTGACTTGTGATCTGCCCACTTGG  
GCTCCCCAAAGTGTGGGATTACAGGCGTGAGCCACTGCACCCGGCCACT  
AATTTTTAAATTGTTAATAAAGACGAGGTCTTGCTATGTTGCCAGTATG  
GTCTTGAACCTCGTGGGCTTAAGTAATCTTCTGCCTCAGCCTCCCAAAGTG  
TTGGGATTACAGGTGTGAGCCACTGAATCTGACATTTTTTAAAGTTTTC  
TTCTCTTTACCAAGTCTTTTTTCCCCTTTCTGCTTTTTTGGGTTGTTTA  
TTTTGATCTCTATCTTGCTAGAACTTTCTGCAGACGTTTAGTAATACTA  
GATTTTTGAGAGTGGGCAACTGGAAAGCTGATTGGAACTCTGAATACAT  
GGGTGAGGCTTGTGGCTGTGAGTGTCTTGTGATGTCCTGGCAAGGC  
CAATGGGTTTGGGACCCCTACTATTAGTATAGGCCTGATTCCTGGGAAA  
GGCTCTTTTGATCTCCTGCCTGGAGGATAAAGGCCTGGCTACCAGCCTTC  
TGTGTGTAATGTGAGGGAGAAGGGCTGGAGTATTCAACATCATGCTGAAT  
CCTTCAATGATCATCTTGTTTTTTAGTAATCTCCTACCTTAACTCTCTGT  
CTTCTGCTAGTATGGGAAAGATGACCTGAAAATCTAACCATTTATTTTC  
CCCCATTAATATCATTTTATGATTATTCAGAAGTTAAATAATTGTCATGC  
TGTCTTCCAAAAGACTGAATCAACTAGCAACAAATAAGAATTTTCTCAC  
AGCTCTGCCAGCATTTTAAAGAATAGCTTTATTGAGCCCAGGAGGTCAA  
GGCTGCAGTGAGCTGTGATTACACCACTCTACCCAGCCTGGGTGACAGA  
GCAAAACCCCTGTCTCAAAAAGAAATTTAAGGAACAGCTTTATTGTTGTA  
AAATAGACATCAATAAACAGAGCACATATTTAAATTGTGCAACTTATAC  
TTTGATATAACCTGTGAAAACATCACCACAATCAAGATAGTGAATATAT  
TTATCACCTCCTGATACAGTTTAGCTCTGTGTCCCCACCTAAGTCTCATG

TTGAATTGTAATCCCCAATGCTGGGGGAGGGGCTTTGTGGGAGGTGATG  
AATTGTGGGGGTGCACTTCCCCCTTGCTGTTCTTGAGATAGTGAATGAGC  
TCTCATGAGCTCCCTTCACTCACTCTCTTTCCTGCTGCCATGTGAGGAT  
GTGCTTGCCTCTTCTTTGCCCTTCTGCCATGATGTGTTTCTGAGTCCTC  
CCTAACCATGCCTCCTGTACAGCTTGCCAGAACTGTGAGTCAGTTAAATCT  
CTTTTCTTCATAAATTACCCAGTCTCAGGTGGCTCTTTATAGCAGTGTGA  
AAAGGAACATAATATACCTCCTAAGTTACCTCAAGCTTCTTCTTAATTCCT  
TCTCCTCCCTTCTTTCATTGCCAAGCAAACAACCACCTGTTTTCTGTAC  
TATAGATTAGTTTACATTTTGTGGGTTTTTTTTTTTTTTTGTAGACAAGGTC  
TCACCTGTTGCCAGGATGGAGTGCAGTGGTGGATCATAGCTCATTGC  
AGCCTTGAACCTCCTAGTTTCAAGTGGTCTCCCACTTCAGCCTCCTGAGT  
ACCTGGGACTACAGGGGTACACCACCACAACCTGGCTTAAAAAATTTTTTA  
AATAAAAAATGGGGTCTTGTATGTTTCTCAGGCTGGTCTCGAACTCCTCG  
CCTCAAGCAGCCCTCCCTCCTTGGCCTCCCAAATGTTGGGATTACAGGC  
ATGAGTCATGACTCCTGGCCTAGTTTACATTTCTAGAGTTTTGTATAAA  
TGGAAACATACAGAATGTATTTTTTTCGGGAGTGGGGGAGTGTCTATT  
TCTTCTTTCTTTTTTCTTTTTTTTTTTTTTTTTTTTGTAGACGGAGTCTCG  
CTCTGTCTGTTGCCAGGCTGGAGTGCAGTGGTGGATCTCGGCTCACCG  
CAAGCTCCACCTCCCGGGTCAAGCAATTCTCCTGCCTCAGCCTCCTGAG  
TAGCTGGGACTACAGGCGCCGCCACCACACCTGGGTAATTTTTTTTGT  
TTTTTGGTAGAGCGGGGTTTACCATGTTAGCCAGGATGGTCTCGATCT  
CCTGACCTCGTGATCTGCCCGCTTCGGCCTCCCTAAGTGTGGGATTACA  
GGCGTGAGCCACCGTGCCCGGCCCAAGTGTCTATTCTTAACCAGCTT  
TCATGCAATCTTTTTTTATTTTACCATCTCTGTGATCCCACTCCCAAAGG  
TACTAGATGTGATTGGTCTTAGGATCAGCTACCATTTGCCCAACTGCT  
TTCCAGCCTTCCAAAAATTTTTTCTTTTTTTCTTAAAGATACTCCTGTG  
TGAGGCTCAGAACTCTTGAATTGCTACTGCAATATGAACCTCGGTGATGT  
GAATGCCAGGGAATTGCCTGATTGATCAAAGAAATGTATCCCTTCTCCC  
TCACTCTGTGCTCTTCTCATTGTGTTTTCCCATCCTTGTGGATTCTGTA  
ATTTAAATATCCCTTTAATGTTATAATATTTAATGGCGTTTGGCGAAAA  
GTACAGAATTAGGTGCAAGAGTGATAGCTGTTATTTTTTTTTTTGGCCTC  
TGAGACTGTTTATATATGCAAGTTATTTAACAGAAAGTTCTGCAGTGACC  
TGAGATGTGAGGGGGTCTGATAGAGTACGTTTGAAGGCAGTTACTGGAA  
AAAAATAATGCCATTTCTGGTTTGTACTTCGGTAAGTTCAGATGACCCAA  
TATATTGTTTACATGTGGCATTGAGTAAAAAGTAGCTTCCCTCCCTTT  
CTTCTTCTTTTCTCCTTCTCCTGCTTCTATAAAGCATCTGCTTGGGAAA  
CTTCTTAGGAGGAGAGCTTGCCAGCCCGTGGGTAATGGAGAGGTCTTGCA  
GAGATAAAGAGATGCTCCCACTCAATGCAGGATGGTGTGGAGGTAAATG  
GGGATACAGTCTGGCATCACTCAGGAATGGGCCTTCTGGCAGGGAAGAGA  
AGGGAGGGGAAAGAGGAAGGGAGTCAAAGATGAATTGCTGAATACGGGGA  
TTCCAGGGCCTGGAGCCAGGAAGAGAACTTTGGGAGGTGTGAACCTGGAG  
GGCATCAGCTGATGAGGAGCAGCCTGAAGTCCGGGGAGGACCTGTTTTTG  
GTGGCCAGGAAGAAAGTGCCTTCCACACACAGGGAGGCCACAAGGCTGAT  
GGGCTGGGGGTGGAAGGACAGCCCTAGGACAGGCTTGGGAAGCAGGCTC  
AGGTAGGGACTGCGAGGTTCTTGTGAGTCTTTTTTCAATCCTGGTCTTAG  
AAAATAGAATCCAAGGCCTCTTGAGAGTGGAAGGTGGGTTGGGAGGAGGG  
CAGATGGGGCTTAGGCCAGGACACCCGTAGAGCTACTGCCAGCTGTCT  
CTCAGGACTCTGCTGAGGTCACTCCAAGGATCATTCTTAGCCTTGCTAG  
ACAGTACTGACAGAGGGAACCGTAGTATCGCACCCACTTCTTCTCTTTC  
AATGAAAGTTTAAAGGTCACCATTTCCTCTGGCAAAGGAAGTTCCACAAA  
TATTCATTTCGGTCTTAGAAAACAGCAAGGTATCAAGCAATTGCAAACT  
TCCTGTGCTGGGGAATTCCCAAGGAAGTAGGGGCAGAGTTCTGGTGGAGA  
CAAAGTGAATCCGAGTGATTAGTCAGTAGCAGTAGCAGTAGCAGTAGCA  
GTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGC  
AGCAGAACCAGAAATTTCCCGCACGTGTCTCAGGCTCTCATTTGCCAACT  
CAGTCTCTAAGTATTTTTATTGGCAGGAAAAATAAAATAGCTATGAGTGA  
AATAATTCATTAGACCTGAGCCTCCATCAATTTGTGTTTAAAGGCCTGA  
CTCTCTTTACCTTTCCCTGGGATGGAAGATGCAATGTTCTCTGATGTCAC  
TGTCAAAAAAGAAAGAACAGTGGGTATATTGTATGCTTGAGTTCCAGCCA  
TTTGTCACAATAGATAGAGATGACTGCCATGTGTGTAGACTTTCTATAGA

FIG. 4 (37 of 61)

91/118

CTGTGTGCTAAACCCGAJCTGCCACTTCCAAGGAGTAGATGAGGAATG.C  
CATGGTTCTGGGGAGCCCTACCCCAATTTGGGGCAGACATTCCAAAGCTC  
ATTTTCTGTGGAGGGGGTTGATGGTTAAAGGACGGCCTGGGAGTAACTCG  
TCTGTACTAGGGCCAGGAGAGTTACATGCTGCTTCCCATGTTATTCATC  
ATTCCTCCATGTGAATAGCTATGGCGTGAGGTCCAAGGTTAGGGCCTTTC  
TACCATAAATGGGGGAATAAAATTCCCCTACCAGCCTGAGAAGTTTCTGT  
TATAAAGAGGCTTTTTTTTTGCGGGGGTGGGGGAGCAAGCCGACTAATGT  
GTTATTTCCATACGGTTTGTTTTAAATGTAGATGTCATATGCAGGAGAG  
GTGGTGTAGTGAGTCACAACGGGATTAGAAGGACCAGTCCGAAAAGCAGA  
AGAGGGTCAAGTTCAGGGCACTGAGGACTACTGCATTAGTGGCGTGAAA  
GGCAGATGGCTGAACAGGAGGGGGACATTACATTGCTTGTTCCTTGAG  
CCTCGATTTCTCATCTAAAAAGAGGGTCATTTATTCACAGAACATTTAT  
TAACTTGTGCCAGGCACCGTGCCAGGAGCTGGACTAAAAATTAAATCCA  
CCCCTGTGAGCTGCTCTGAAGGCTAAAATATGAAGTATGTAAAAGTAACC  
AAGTGCTGTACACATGCAGCTATTCAATGACTGTGTGGGCATTGCGGCAG  
ATTTTAATTTTCTTTTTTATTTCTTTCTTTAGTGAGAGGTGTTGGTTG  
TTATTATTGTCGTCGCTGTAACGTCTATTTCACTTGCTTTTTTGTGTC  
TCCAGCCCATTCCAGGGCTGTCATCTAAGACACTTCTTATCACCTAAATA  
ACCGGGGAGGCAAAGCGCTTCTTAAGAGATGGATCCAGAAGAACAATGC  
TGGTTTTCTGTAGAAAAAGGGGCTGTGGGAAGTAGAGATAAGAAGGGAAT  
TGGCCAAAGATGAATGTACAGAGCCTTATTTTTTTTTTATAACACAGCAAG  
ATTAGATACAAAACAGGACAATAGCATCATCTGTTTTTATAACTGGAAAG  
GACCTCACTTTACAGGTGGGGAAGAATAGAGTGGAGAAGTGAAGAGAATG  
GTCACAGAGTCAATCAGCATGTCTGCGTCAAAGCTGGGATTCCCAATTCA  
GGGCTCTTACTACAGTGACGTATGGCTAATATTTTGGCATTGTTTCGGGG  
AAAAGCTGAAGCCCTGATGGTGTACGTCACTCTTGAGATAGTCTGTAGTC  
CAGCAGGGAGGAAAGCAAGGAAGGGAGGTGGAGGCAGCATTTTTTGGGTGT  
AACATTTCGTTCTTGTTTTGTGGCCAAATCATAGTGTGATTGGGACAAGC  
CACTGCCTTTCTCTGAGCCTCCACTTTCTTTTCTTCTTAAGAGGGAGGG  
AATAGTAGAGTAAAGTAGTCATTTTATCAAACACCTGCTATTTTGGAGC  
CATATTGCAAGTGGGTGGGGGTTGAACACTTGGCTTTATTACCCATAGG  
ATTAAATCCAACCTCGATACTGTGGCATTCCCAAACCTCCAGTCTAATCTT  
CTTCTCCATCAGCCATGCCCCACGACACCCTGGTCATATCTGATGTTGCC  
CCTTGCACTTGCCCCCTCCTTATCTTTGCTTTCTGACCTACCATATGGCT  
ATTGTTGAAATTCTCATTTTCCAGGGCCTTGCTTAAATATCATCTCATC  
CATTAAACTTTCTTGAACCTCCCCCTTGCCCTGTTCTCCTCCTAATGTCTC  
AAGCCAGAATTTATTTCTTTTGTGGCCAAGGGACTGGGTTTGTGACCTC  
TCTCACGAGACTTAATATTGAGACCAAACGTCTTTAGACCTCACCAGCCA  
GAGATGAGCATCTATGGAATGCAGGCTTTTGCCTGGACTTGCTGATGC  
AGGGCCTCTGCCTTCTCCAGGGCCTCTCCTGCTGTTTTAGGAATTTCCC  
TCATGGCACAGTCCATGAGCTCAGGGTCAAGTTCATACATGTTTTTACTT  
CTTCTACTCTGCAAATGGTCTTCTTGAACCTCTGAGGGTCTTAAAGCTGCT  
CTGCAGTTTGTGGGGTGAGTAGAAAGGGGCTTTCAAAGTTGTGCTGTTG  
TTTCCCACCCCAATAGCATGAAACACAAAGATGCTTACAAATAGCTGCCT  
TGCTTTCTAGTCCCACTTCTCTCTCCTGAGGCTTTAAACAAGTCCCCCT  
AGGTTGAGCTGGACTGGAGTTGTATCCTATCTTCATTATCTGTCTACTCT  
CTTCTGCTCTCTAGAGAAGATATTATATATGTGTGTATGTATGTGTAA  
TATATAATATCCATATATAGAACATATATTGTTATATTTACATATACATA  
CATAACATATGCATGTATTTCATATATACATATGTAGTATCAAAGTTGGAA  
TTAAACTGTATATTTTGTAAATTTGCTTTTATTTGCATCTATCACTGTAA  
ATGAATATTTATCCATACCGTAAGATATTCTTCAATGTATTTTTTTTTTT  
TTTGAACAGGGTCTTGCTTTGTTGCCAGGCTGGAGTGCAATGACCCGA  
TCTTGGGTCACTGCAGCCTTGACCTCCCCGGCTCAAGTGATCTTCCCACC  
TTAGCCCTCTGAGTAGCTGGGACTAAAGGTGTGTGCCTCCACACCCAGCT  
TTTAAATTTTTTTTGTATTTTTTTTTTAAAGACAGGGTTTTGCCACATTG  
CCCAAGCTGGTCTTGAGCTCCTGGGTCCAAGCAATCCTCCCACTTTGGCC  
TCCCAAAGTGCTAAGATTACAAGCATGAGCCACCACACCTGGCCTCAATG  
TAATTTTTAATGGCTGTATAGTATTCATCATGTGGTTGTACCCAAAT  
ATTTAACCAGTCCCCAGTTTATTTCAATTTTTTTTTTACTATTTTGAATA  
TGTTTTAGTAAATACCCACAAAATATGTACAATGGCTGGGCTTAGTGGCT

CACCCCTGTAATCCCAA<sub>1</sub>ACTTTGGGAGTCTGAGGCAGGTGGGT<sub>2</sub>CACCT<sub>3</sub>  
AGGT<sub>4</sub>CAGGAGTTCGAGACCATCTTGGTTAACATGGTGAAACCCCGTCTCT  
ACCAAAAATACAAAAATTAGCCGGGTGTGGTGGCACACACCTGTAATCGC  
AGCTACTT<sub>5</sub>GGGAGGCTGAAGTAGGAAAATCACTTGAACCTAGGAGGCGGA  
GGTTGCAGTGAGCCGAGATCACACTACTGTACTCCAGCATGGGCAACAGT  
GAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAGTACAATTTGTTG  
TACCTCCCTGATTATTTCTTTAAGTAGAATTTCTTATAATTTTTTTTA  
TAAGTAAAATTTGAATCAAGGGAGAAGCACCTGGAGTCCTTCAGATACC  
TATTGCCAAACTGAACTTTTCTGTTCAGGTTTACTACATT<sub>6</sub>CAGCCTGAC  
TCAGGGTTTGGGGAGTAGAGGAGGGGTGGAGGCAGAGGGCCTCTCCCTG  
TCCCCACAGACCTCCCTTGGTGAGGTCCAAGTCTGGACAGGTGGAGTGTG  
GCATTGCACCGTCAGGTCTGTCTCCTGTAATTTCCCTAAATCCATCCAG  
TGGAGCCTCATTTGTTCAAGTCTTTTTTTTTTTTTTTTTTTTAACTCCC  
CTGAAGACGGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCAGTGGCACGA  
TCTTGACTCATTTCAACCTCTGCCTCCAGGTTCAAGTAATTCTCCTGCC  
TCAGCCTCCTGAGTAGCTGGCACTACAGGCGTGTACCATCACGCCCCGGCT  
AATTTTTTTTGTATTTTGTAGTAGAGACGGGGTTTACCATGTTGGCCAG  
GCTGGTCTCGAACTCCTAACCTTGTGATCTACCCGCTCTGCCTCCCAA  
GTGCTGGGCTTACAGGTGTGAGCCACAGGCTGGCCTCAAGTCTATTTT  
TTAACTCCAGGAGGCTGGTATTCAGAGGGATTAGGGCTGGCAGAGGGC  
CTCAAAGCTTTCAAGGCTGGGGAATAGGCTGCAGCCTGGTTCAGGGTAA  
CCCAAGTGATTTTGGTTCCAAAGGGACAGGAAAAAAGTGATTGATATGG  
AAGTTGTCAAAGTGCAACTGTCAAGACATTAAAAATGTAACCCTTTTAC  
TAATATACAGTAGACTTGTGTTAAATATTTAACTGATTGTAAAAGGAAAA  
AACCAGACGCAGTTTTCCCTACCATACTGTCAACACCTCAACACTGAG  
TTCTTCTGTGACCTCTAGTCACCGAAATGCTTGGGGATTTCTCCCACCAC  
TAGTCTCCAGCAGCCGACACCAGTTGGGTGTCTAATTCACCTCAACAC  
TATCTACCTGGAGTTAGCGTTAGATCCACAGGTTGAGGGCTCAGTCTCA  
CAAGACTGCCTCCCACTTCAGGTGCCAGTTACAAGTGGTAGGTTGTCAAC  
TATGCTTCTGACTGATGGCTATAAATCTGGGTTTGTCTCCCTCGGGTTCC  
GTGAATTTGCTAGAGCAGCTCACAGAACTCAGGAAAACTTAAGTTTAC  
CAGTTTATTCTAAAAGATATTACAAAGGATACAGATGAACACCAGATGAA  
GAGATGCGCAGAGCAAAGCATGTGAGAAGGGGTGTGGAGCTTCCATGCCC  
CTCTGGGGCACCACCCTCCAGGAACCTTCATGTGTCCAGCTATCTGGGAG  
CCCTTCCAAACCCTGTCTTTTTGGGTTTTTAAGAGTGGCTTTATTACAT  
ACACATGATTGACCGAACCATTGGCCATTGGTGACTGACACAACCTTCAG  
CCCCCTCACTCCCTCCAGTGGTTGGGGAGTGGGGCTAACAGTCTCAAGTC  
TCCAATCCTGCCTTGGTCTTCTGTGACAAACCCCATCATGAAGCTACT  
GCATTGGGGCTGCCAGCCAGTCAGTCATCTATTAGCATGCAAAAGACACTC  
TTATTATTCCAGAGATTCCAAGGGTTTTTAAAGCTGTATGTCAGGAAAC  
AGGAGATGAAGAACAAATATATATTTCAACATCACACTCGTTGGGGGA  
ATTGACAGGATAGCAAACTGATTAAAGGAGGATAGGAGAGACTGAGATA  
TATATTTCCATATATATATATAGAGAGAGAGAGATATTTCCATATATA  
TATATAGATCTAGAGAGAGAGAGATAGAGAGAGAAGAGTCTTTCC  
>Contig51  
ACACATTTGGGGGAGCAGTTCCGGAGGTACAGCCCCGACAGGAGATGTGA  
GAAGATCGTGGTTANTGTTCCCTGGTCCAGAACCCCTCCAAGTGGGCTT  
AAGTAGGAAGGGTGGTGAGCGGCAGGTAAACACACGTCAAAGGCAGTCTT  
CCTCTCTGAGGGAAAACTTGTATAAGCATTGCAATCAATGGGCCTCTT  
TAATTATGTGCCAGTGGCAAGAGCGGGTGCTGAACCCAGGGGCTGCCTC  
AATCCGGGGCCTTTGAGGCAGAATAAAGTGGTCTCAGGTTGTTGGCATT  
CCTTGCCCTTCCACCCGAAGCAGACACAAATCCTCTCTGGAGGCAAGTTC  
CCCAATTAGCCAGTACAACCTCCACAGACTAAGATCAATCATGTACAAG  
CTCACAGACAAAGGTCACCAACACACAGAGCAATAAACAAATTCATGAG  
TGACGTGAATGAGAATAAACAGAAACAATAACCACCAGCTGGGATGCTCT  
AAGTCTTCAGCTGTTAGAATTCCTGAATATAGAATAAACTGCCACAATG  
GCAAACTAGCATCTAGTACTTACTGTGTGCTGGGTTCTAAGAAATTTGCA  
CATTGTGCCAGATACCGACTCAGCTTCACACTCACCTCCTACTGTGCCC  
TCTTAATTTGCACTAGATTAAAGGTAGAAAGGAAGAGGCAGCTATTCTG  
TTCTTGGCTGTGCCTCTGGCAGCACATGCAAAATGGGCAGTAACAGTGGC

AGTCACAGGTAAGTAGC TTCTCACAGT JGGAGTTAAAGGCATGGGA  
GAGACGAGCAAGGTTTCTTAAAGGGACAGTGGCCAGTAAATGACCAGGGGC  
TACTGGAGTGGCTGCGATGGCTCTGTGGAAGCTCAGAGGAGCCTTGGGTCC  
TGCAGGTGCAGTAGCAGCTTTCTGTAGTTCTGTATCTCTGGGTCCCACAA  
TCTTCCCCGTTTTTGTCTCCTCCACTTCTAATTTTGTAACTGACTTCCCTG  
TGTGTACTTCTCTCTCTGATTGAAATAGCCAGACTGGTTTCTGTTTCTG  
ATAAGACATTGTCTGGTACGAACACAGTAACTCATTTAATCCGATATCTC  
TATGAAGGAGGTACAATAATTATTCCTATTTTACAGATGAGGAAACACAG  
CAGAAAAATAAAGTCAATTGTCTAAGGTTGCACATTTAGTCAAGGGAAGG  
GTTGATATAACATATAATTATTTAGAAAACATCTAAGGAAATAAAAGGCA  
TAATTTAAAAATAAACTAGGCAGGTTTAAAAAATGAAGTAATCTATAA  
GTAAAAAAGTATAATTGTTGAAATACATATCTTAGTGATGGGTAAATA  
GCTGAAGAAATGATTAATGAAGTGAAGGTAGTTCTGAGGAAATCAGAAT  
TCAGCATAGATAGAAAAATGGGAATTTACAAAAGTACACAGGAATTATA  
AAAGAGGTTAAATTATAGGGAGGGTAGAATGAGAATTAACATTGGTCTAA  
CTGGAATTTTGAAGAAGAGAATAGAGAGAATGAACAAGGCAATATTTAA  
AGAGGTGGCTGAGAATTTTTCAGAACCAACACAACTATGACTTTACCAG  
TAGAGAAAAAATGTACACTGAGGAGGATAAATAAATATACTATGAACAA  
ATTGTAATAATAACTCAACAAAGACAAAGAGAAGATCTTAAATCAGC  
AAAAAAGAAAGTCAGACTTAGAAAGAAATGACAATGGCAGACTACTCAA  
CAACAACATGGAACCAAATTCAGTGAACAGTATTTTCAAATGCATA  
TTTAATCTATCTTTGAAGAATAAGGGTGAAAAGGGTGAAAATTGCTGCCT  
TATACAAAATATCAACATTAACAAAAAGTAATGAAGGTAATATAAAAAATG  
TTTTCAAATAAACAAAACCTGAGAGAGTTTACCACCAACAGCATTCTTA  
AATGGACTTTTAAATGCAGTTTTTAGGAAGAAGGAAAAACAATTCCTAAGG  
AAGGTCTGAGATGCAAAAAGGAATTATGAACAAAGAAATTGTTAAATTA  
TAGGTGAATTAAAAAAAGTGCCTGCATAAATGATAAATGACAATGATG  
CTATTAATAATGAGTTGATAAGGATAAAGAAAAGGACAGAATTAAAATAC  
TAGAAAAACAAGCATGCTGGAAAGGATTAGGAATTAATTTGAAGGTTAAAG  
TTCTAGGGTCTTCTATCCTTCTAGAGGGGAGTCAATATATTAATTTTTG  
ACCGTCACTTACACAGTGAAAAACTTTAAGGATAACCATAAAAAATAGA  
AATAGAGATATAACTTCTGAAACAGTCAAGGGAAAAATATGGAATAAGA  
AAACTGACCAAAAAACATCTCAGTCAATCAAAAAAAGAAAAAAGAAA  
GAAAAGGTTCCGAAGGAGAAAATCAAAGCATAGAAAAAGCGGGACAAATA  
GAAGTGGAAGAAAAGGTTAGAGAAAACAGGTCCAGAAATATCACTGAT  
GCACTAAATCACCATTAAAGATGAAAACAAATGAACAACATCAAAAAAT  
TCTAGTGACTGTAGTAGTGCTGATCAGAATAGGCTCTAAGATAAGATGCA  
TTATTGTGAGTCAACTTGTGATGATGAAAGGTTTAATTCACCAGAAAGAC  
ACAATTATAAACTTGTAATCAAATAGTTTTATTTTATTTACTTTATTTAT  
TTATTTTTTTTGTAGACAGGATCTTGTCTGTGCTCAGGCTGGAGTGCAG  
TGGCTTGATCTCAGCTCACTGCAGCCTCCACCTCTTGAGGCTCAAGCTTT  
CTTCCTGCCTTAGCCTCATGAGTAGCTGGGTCCACAGGCACACACCACCA  
AGCCCTGCTAATTTTTGTATTTTTGTAGAGATGGGGTTTTACCATGTTA  
CCAGGCTGGTCTCAAACCTCCTGGGCTCAAGCGATCTGCCCCCTCGGCTT  
CCCAAAGTGTGGGATTATAGGCGTGAGCCACGGTGCCTGGCCTCAAATA  
ACTATTTAAGTGAAACAAAACCTAGTATGGCACTAATGAAAAATGTATAAA  
TCCATAATCGCAGAGGGATTTCAACTTACTTCTTTGATTATGTAAAGGT  
CAAACAGACAAAAGACAATGACAAAACCTTAATGCAATGAACACTTTTGAT  
TTAATGAACATATATTGGATATGTACCAAGAATTAGAGAATACATACTA  
GTTTTGAGTTTATGCAGAACATTTACAAAAATTTAGTGGAAGCCTAAATT  
ATAAAAAGTTGCTGTCTACGTAGAATAACACACAAACCCCTGAGTCCGGAA  
TTCAAAGCCCTCCACACTCTCCTCTACCTTTGCATCTTTATCCTCCACCA  
CACTGCAGTGCATACTCTGGGCTACTACTCACTGTTCTTGATTCAAATTC  
CATGTTCTGTCTCAGCTCAAATCATTCTCTCTGCTGGAATAACTACTTCAT  
ACATATTCTGCTATTGAATTCTTGTCTTAGCACCCCATCTACTCCAAGAC  
GATGTCCAGTTGGGGTTACTCCCTGTCCCATTTTCTTTGATTACACTTTT  
TTTTCTACTTCCATTATATTATTGATCAGATCTGTGCCACAGTTTTTTGA  
CTTTGTGCTGCTTTTACTCTTTCTAGACCCTGAGAGCTCCTGAAGGGT  
TGGGTCAATTTCTTTTATTTGCTCATTCTCATGGCACAGTGAGTGCTT  
AATAAATGGCTATTGACTGAAATTAACTGTATCTAATGGACATATTCC

FIG. 4 (40 of 61)

94/118

ACTTCTGGGCCATTTCATCTTTCTTTCTATTGGAACCAGGAGATGGGGAA  
CCATAACAAAGGTAAGGTTGTGCCATGTGAAAGACATGGAACCTTCCCC  
TGAGGGCCAAAAAGAGCAGGGAAAGGTGCAAAGACAAAATCTTCCATTT  
TTAAACAATGTAAGAATGTGGTCCACCTCATGCTCAGGTGGGACTTTATC  
ATGACGTTATTTTTGGGGACTTATAGCTGCATCATTTACCCCATATACAT  
TTACCTTTAGTGTAGGGAACTGAGGACAGGAATTTTGTGTGATGCAGACTC  
TTGCTAATGAGGCTAACACTTGGAGAATTTTATCATGCATTCAAGAAGC  
TTGTTTACATTTCTTCATTAATACTTTAGTTGGTGGTTTAGCTTTAGTT  
GTAGGCTTATCAGATATTTGGAGATATCTTCATAAACGATGGCTTTGGTT  
TTAGAAGAGTTATTCTGAAGCTACTATTTCTGGCAATAATCAAACAGCAT  
GGCCATTTGTTTTGTAAGGCCTTTCCTAAAATATGACGGTAAAATCTACG  
TGTGGA AAAATGCTTATTCTTCTGTCCTCTATAAATGTGAATCTAGTTTG  
TCTTCAAATGAAATCAAGTGATTAAATGTAGTTTCTAAGAAGATAAA  
TGGAGCAAAGCACTCTGTGTTTACAGTGTGGAAATCACTCATCCCTCA  
TAAACTGTCCCAACTGATCCTGACTCACATGAATGAATTAATAAAGAG  
TTAATAACATCAATTTACATTTTTAAAGACACTTCCCATGTTTTAGACT  
ATTGGTTGGA AAAAGCTGGTAGGTGTACAATTTGTGGAGAGTTGGCTGTTT  
TTGTCTGTCGTTGTTTGACGTATTTCAAAGCCATATCTAATTTTGTGCA  
GAATGGTCTGAATTTACAAAAATGTTGAGTTGTGTAGTGTGGAGAAGTA  
CGGAGCCATTTACTGAAAGGCTGGGGGGAATGACGAGACCCTGAGATAA  
GGCAGTAGTGGTGC GAACAGAGTGGAAAGGGAGGTAGTTGAGATATGTTCA  
GAGTAGAATCAGAATGGACATAGTGAACAACTGGATGCAGGTGGGGGCTG  
AGGAAGCAAAGTTGAGGATAATTCTGAGACTTCTAGGTTGATCCACTGAA  
GTTACATTATTCAACACCACAAGGAACTAGGGGAATGAGAAGGCATACT  
GGTTTGCTTTGGAGTGGAAAGGGCAGTGATGTAAGAGGAGTTAATGAGTTA  
AAGTTTGGATATGCCTGAACCTCAATTTGATATGTGCATCTGATATACCC  
TTGGGGTGACCCTCCAGGCAATGGTTGAACATGTGTATTTCTTAGTAAT  
GATAGGCATCACAGACTCACATCAGTAAGGAAGCAACAGCAAATGATT  
GGAGCATATACCTGGAACCTCAGTACCCTATGACTGGAGCAAGTCTCTGTC  
AGTGAAATGAGGATAAGAAGAATCTTGACCTTGTGGAATATGTTGTTAGG  
AATATATGTGATGAACAACATAGGATACTTCCTACAGGGCTCCACATGTA  
GTAAGGGCTTTATAAATGCTTGATAAATATTATTGTTGTAATTTATTTCC  
AAAGTAAGATGCCACTGGAGGAATCTTTGGAACCCAAATTAATAACAAAT  
AGGACTGGATGCAATGGCTCACACCTGTAATCCCAGCACTTTGGAAGGCC  
AAGGCAGGAGGATCTCTTGAGCCCAGAAATCAAGACCAGCCTGGGTGAC  
ACAGGGAGACCTTGATCTATGAAGAATTAAAAAAATTAACCAGATGTG  
GTGTTGCACGCCCTATAGTCCCTGCTGCTTGAGAGGCTGAGGTGGGAGGAT  
TGCTTGAGCCCCATGAGGTTGAGGCTGCAGTGAGCCATAATTGTGCCACCA  
CACTCCAGACTGGGTGACAGAGTGAGACCCTATCTCAAATAAAATAAATAA  
ATAAATAAATAAATAAGTACAAACCAGCAAACACTAATCCTTTCTAGAGA  
TTATTGAACTCTGGAGGGCAGATCTGAATGGAGCCAGCAGAGGGACCTAT  
GGAGATCAGCCTGGCCCTGGACAGCACCAGGCAATGGGGTTGCTAGAGAG  
GTAATGGGGTTGAACAGGGTTTAAGCCATGAGGTCTCAAGAATCCGTGAA  
GACTCAGACTAATTTTTTTTTTTTGCATGAGGATTAGGTGTTCTTAGGA  
ATTTCAATGAGAGCAGGGTTAATGAAGGAATGCAGGGTAGGAGAGCTGAG  
GGAAGGCATCTGAGAGAGCCTGGCTTATGAATGGCTGCGTCAGTATGGCT  
CACCTGCTTTCCTTGTATCTACTTAGCAGATGATCCACCCCAGGCCTCC  
AGGGCCAAGGTCATTTCCACATAGTCATGGGCCCTTGAGGGCCTGGAGCA  
GTGTAAGGAAGACAGAGTCTTAAGAAATTGCATTAACAGTCATGGTGCTT  
GGCAAGTGTCGTCATCCTATGCCAAGCCTGATCTGAAGGGGTGCATGCTC  
ATAGGTAGCTGCTGCCAAGATTACAGCAGCTTCTTCAATCCCAGATCCA  
TGCTCTCCTATATTCAATTTTCCAGGGGTTCTGTCCTTCGACAGTGATG  
AGATGCAGAATGACTTATTGAGTTATTCTCCTGATAGTTGCCAACTTTTC  
CAAATGACAATGGGGCATGGAGCTTGAGAGTGGAAATGAGGCCCTAGGGA  
TAGCGTGCTTAGGAAAACACTCCAGCCTGATGTAATCTGGGGGTACAA  
TGGCATTTCATCATCAAGACTGATGTAAAGGGTGACTAGCAGTGAGTTG  
GGGGTGACTCGCACTGGGGCTAGGTTTCTGATTCTGCCTAATCCAGACAG  
AGCAGAAGCACTAGTGGGCTGGTAGAGGGCCTCCAGGGCCTCACTTAATG  
TCCTGGAAAAACAGCTCCAGATTGTTGGTTTACGTTCTGAGGACAAGCTT  
GGGTACTACAGGATAGAGAGAGTGGTGGGAGATGCCGTGGCCTGCCCTGC

TGATGCCTGCCCTGCCATTCTGCGTGTGATGTCTCTGGGGCATCTTGCC  
TTCCCTGCCAGACCTGTAGTTCAGCTGAGGGCATGTGGAGGCCAAATGG  
CTTCTTAGAGTGTACTTTCTTGAACAGCTCTGCTGGGAGAACTGGAGG  
AGCTAGCTAGTCACGGTAACTGCAGCAGTCAAAGGATCGTCCCGGTGGAG  
GTGGGGTGGAAAGGTAGAGAAAGAGAACATATAGCGTTTTCTTGGAGAT  
GTGTGGGCATGTCTAGAGGAAATACCCAATTCTGAGCCTTGAGCCCTC  
CAGGAAACCTTGGAATATTAGGTTAGTCATCCCCAAGGAAGTCTAAGAAT  
TCTGGTCTCACCCTCTCCTTTAATTCCCACAATGATCCTACATGATATT  
AAGGAACACGGGCGAGTAACCCTCCAAGCAATGGATGTGGTGGTGAAGTT  
TGACCTCATGATGGAGCGGAGTTGGTTTGAAACCTAAGAATTTAATTTA  
TTGTTTCAAACCTGTTCTCCACTCAGCGTTATTAAAGCATAACATAATTGAC  
ACATAAAAATTGTATATGTCTACGGTGTACAATGTGATGTTTCGATCTAT  
GTATACATTGTGAAATGATTACAACAAGCTAAATAACATACCCATTCTC  
GTGTTTCAAAGGAATTAACTCAAGCACAAAAGAGAGGTGCTGTTGAAGA  
GTAGGGCTGCTCTATCTAAGTAGTATGTCTGGGGTTGCTCTGGATCAGGG  
TCCTTTTGTGCTAGTAATAAACAGCCCTTCTGGGGCTGCTCεACTTTCC  
CCACATTTTCTTCTGGAGCCTCCCTAAGAATTAGGACATGGCCACTTTCT  
CTGCATAGGCTTCTACTTCAACAAGGACAGGGCTTGTGCTGCCCCATGC  
CACTTGAGTGTCCCTACAGCACAGAGCTGAGTGCACACTGGCTGAGTGAG  
GAAATCCCCCAGATTAACTTGGTTCTAAGCATCATGGCTGTATTTTACA  
CGTATATGAATTACAAATTACAGCATAGTCTGAATAAGGATTTTGTGCTA  
CAACTGGAATCCAGATTATGCAAAATTGGATAGTATAATATTGAAATTCC  
TAGGACTTTTTATTAGTTTTTAAAAAATTATACAAGCTTAGAGTAAGAAAT  
TAAACAGTGCAAAAGAATTCACTGTGAAAAGTAAATGCTCTGTCTCTGC  
TGAGAGACAGATATTGCAGCCAGATACTACTGGGGTCAATAGTTTCCTT  
TAAGCATGCCATTTTGATGGTTTATGGGACTTACAGCTCAAGAAGCTTGA  
CACTAGGGTTGATCTCAGAAAATCATTGTTGCAGGTATTAGATATGACCG  
TCTCATAAAGATACACACACAGACACAGCGATTGGAGATATTCAGTGGGG  
CTTATGGGCTGCTTGTCTCTCTGCTCTGTGCCTAAGTTGGGCTCAGAGT  
AGCCTGGCATCGGCTGTGGGGAGAATGCTGGCATGGGGTTAGCAGGAGCC  
CACTTAACATGTCTAAGCCACCTGGAAGAGTCTTCAAGGAGACCAGAC  
TCCAGAGGCCCTAAGGAAGGAAGGACTTTTGCCCGTTTTTAGGTATTCTA  
GTCCAGAGTTTAGGGAGGAATGGTTTGGCTTTGGGTCGTGTGCCCTTT  
ACCGAGTGGGATGGGATGTGCCATGAGCTGTTGAGCTGGCTCTTGGAGA  
AGACAGCAAAAGCGGGAATAAGAGGTGAGGAAGCTGTGTGGTTGTAGGAA  
ATCCCAGCAGAGGGCCTGGGGGTCAAAAGTGGTCATGGTAGTGACGGTGG  
AGGCTGAGGTGGTAGAAAATCAGAGGACAAACCCCATGGGCTGCTGGTGA  
TCTGACCGAGCTCCTATGCTCTCCTGGTTCATTTTAGGCTCTGTAGCAGC  
AGATGATTGGCTGGTGTGAGAGCAGTGCACCTGCCATATCAGGCAATCCA  
AGACAAGTCCAAGCTACGCTGGGAGGAAACCTGAAGGCAGCAGCAGGTAG  
ACTGGCTGAAGACAGACAGGCAGGCAACTTGTCAATCAGATTTGTGTTTT  
TAAGGACTTTTAACTGGGGAGCCCTCCATGACAGATCAGATGAGAGAGGA  
ATCTGGGTCCGCCATGTGTCAAGCTACCAGAGGGTCCCATCGGTGCTTG  
GATCTTCTTTGAAGCTGGGTCTGAGGTTTGCAGGTAGAGGGTGAGCTGGT  
CAGAGGGACCTATTGCAGAGCTAACCAACACCTTCCAGGAATGCAAGCA  
CAAGCACCCACCGCGGGCAGGCGGGCAGGCACTTCTCCTTTTGCCACCA  
GGACCTCACAGAGGCTGATCTGGCTCTGTGAGGTGGGAAAATGGGTGTA  
CTTAGTACATAGAGATAAAAGGCTTAGGAGGCCCTCCATCCTGTGACCC  
TGTCCCCAGACCACAGGTGCCCGCAGGTGCTGCTATTTCAAGGCTGGGCC  
TCAGTGCAAGCTTGTGGTTTCTTGCCACCTGTGATGTCTCTCCACTAAT  
GAAGGGGCTCTCCATCCTCTGTCTGCCTCTAGCAAGTGGAGGCTCTGGGC  
CCTGGGCAAGACACAGGGGGAAATGCCATCTGTTATCCAAATATATTTCA  
ATGTGACAGGAAAGCTGTCTTTAGAGCACAGC

>Contig52

GCATGTGCTCTACATTGATCCCAGGAGTTTGAGACAACATTGCAAGACTG  
GGCAACAAGCAAGACTCTGTCTCTACAAAAAATAAAAAAATTAGTTGGG  
CATGGTGGTACATGCCTGTGGTCCCAGCTACTCCTAAGTTGAAGAGGGAG  
AATGCTTGAGGCCAGGAGTTCAAGGCTGCAGTGAGCTATGATCACACCA  
CTGCACCTCTANCCTGGGTGACAGAGCAAGACCCTGTCTCTAAAATAATAA  
TCGTAATACATTTTTTTTAAAGTAAAACAAAAAAGGTACACTTTCTCA

TACCAAAATAAATTCCAAATAAATTAAAGGCTTAAACATGAGAAAAGTTAA  
ACCATAAAATTACTAGAAGAAAATAAAAGCAAATATTTAGATAATCCTGG  
GGATAAATTTCTTTGGAATGAATTTCTTAAAGATGAATCTCTAAAAGTGA  
AATTCAGGGTTCAAAGGTCTTTTCTTTGTCCTTTTCTTTCCCTTCCCT  
CTCCCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT  
TTCTTTCTTTATCTTTCTTTCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC  
TGGTTGCTTGCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT  
TTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT  
CTTTCTTTCTTTCTACTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT  
TTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT  
TCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT  
AACAGTCGCATGAACATGGCTCACAGCAGCCTTGACCTCCTGGGTCAAG  
CAATTCTCCTGCCTCAGTCTCTCAAGTAGCTGAGACCACAGGCACCCACC  
ACCAAACCTGGCTAATTTTGTATTTTGTATTTTGTATTTTGTATTTTGT  
ATTGGCCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCTGCCTGCCTT  
GGCCTTCTGAAGTGCTGGGATTACAGGCTGGGCCTCTACGCGGGGCCGAG  
ACTACCTCTCTTTTAACTGGATCTCTGAGCTCTGGGCAGAGCCCACCCTG  
AATCCTGGTCTCCAAAAGGGGAAAATTATTAGGAGGCTAGACCATATGAT  
GCTTTTACAGTGCACCTTAAAAAAGTTTGTTTTTTTTTTAAAAGACATT  
TCTACATGTCTAAACTACAATCTTCTTGAACCCCAAGAGTAGCTTCTG  
TTGCAATAGCTAGTCAAAAATATAATAGTCAAAAAAATCAGGTAACACAA  
CACAAACGCAAGCAGTTTAAAGAGCTGAAATGAACCTGTCTGTTTACACTC  
TAGGGATTCCATAAGGAAAAATAGAAGTTTCTCCCTAAAAGGGAGCCTGG  
CACCTTCTCCATTTTCTTTAAGGAACCCCAAGGCTATTATAAACTATTTTA  
GGGCTCTCATGCAGCAGACGGTGCAAGAGAAAGGAGAGACAGCAGAAGTA  
AATGAAGAAAAACAGAATCCAGTCAACAGAGAAGAAAAAACTTTTGCTCA  
AAAAAAGGCAAGTTCTTAGGAAAGAAAAAATCATGAGGGCTATTTAA  
ATACAAAGACGCATACATACACATGCACACATCTTGGATGTTAGCTTTTA  
ATTAAGCTGACTTTTAACTATTGAGGTCTTTTAAATAAATCTTTTAAAA  
TCTTATTACGATATTTTCACTAGGACAAATTGCTGCTATTTTCAGCATTAC  
CAAGTATCAAAACCAGAAAAAGGCTTGATTTAGGAACCAACCCAGGCTGTC  
GTGGTAGGAAAAAAGGCAGAACGTTAGCTATGGAACCCACAGCATGGGGC  
AACAGCCATTGCTCTTTTCACTATGGCCTGGCTAGCAAAAAGGTGGCCTTG  
TTATGTAAATAAAGCCCGTTTGGTGGTCAAAATGAAACATCTTTTCTTT  
TTTTTTTTCTTTTGTGGCCGTTTTTTTCCCCCACCATACCACGTTTGTGT  
GTGTGGGAGGGTGGGAATTTAGCCACTTCAAGGCTCATTCCCCATAAT  
TTGGAAATTTCTTTTGGATTGATCAAGTCAGATAGAGTAGGTCAAACCC  
AATGGGAAAAAGACTGAAACAGCAATAAAAAACAGAAACAAACAGTTAAGC  
AAAATGAATGATCACACAACCTTATATGATTACTGAGTGCTCTAATGGTAA  
GGAGAAATTAAGACAGCTGGTTGTTAACTTTAGCCAAGACAAAACCC  
AATTCAGCTACTTACCTAGGGTTGGGTCTCAGGCTGAAGACCGCTCACTA  
CCGTTCTAGAAGCAAGAAATAAACTTGAACCTCGTCTTACCTGTGTAGCA  
GGACAAGCCGACAGCAAAATCCCTCAGACACCAATTAAGAAGGAAGGG  
CTTTATTGGGCCTGGAGCTGCGGCAAGACTCACGTCTCCAACAACCGAGC  
TCCCCGAGTGTGCAATTCCTGTCCCTTTTAAAGGGCTCAAACTCTAAGGC  
GGTCCACATGAGAGAGTCGTGATAGATTGAGCAAGCAGGGGGTATGTGAC  
TGGGGGCTGCATGCACCTGTAGTTAGAATGGAACAGAACATGACAGGGAT  
CTTCACAGTGCTTTTCTTATGCAAATAACCGATTAGATCAGGGGTCGATC  
TTTACCAGGCCAGGGTGTGTACCGGGCTGTCTGCTTGTGGATTTCATT  
TCTGCCTTTTGTATTACTTCTTTCTTTTGGAGGCAGAAATTGGGCATAA  
GACAATATGAGGGGTGGTCTCCTCTCTTACCTGCGGGGAGTGAGCTCAA  
CTCCTTAAAGGAGTTACCTGCCTTCCATCATCAGGGAAGCAGGAAATCTT  
GCCTTCTTTGTTGGAAGCAAGTAAACTCAAACAAACAAAGAAAAAAC  
AGGGAGTTGTACAGCAAAATAAACTTTTGTATTTTGAACAAATTTTGGGAG  
ATCAGGAATTCTCTGAAGGAGATGCTTTCAGACCTCAGCAAAATTGTCTG  
TTGGTTTGAAGCATAAAGTTAGCTCATGCTGGTACCAAACACCAAGTAGGA  
GATTTGTCAAAGGTAAGAGGCATCTCCACTCAGAATCCCTTCGTGGTTAC  
CAACATGTGAACCTTGGAAATCTGAGACAGGTCTCAGTTAATTTAGAAAG  
TTTATTTTGGACGGTTGAGGACACCCACCCATGACAGAGCATCAGGAGG  
TCCTGACCACATGTGCTCAGGGTGGTCTGAGCACAGCTTGGTTTTACACA

TTT TAGGGAGACATGAGACATCAGTGAATATATGTAAGATGTACACTGGT  
TCCCTCCAGAAAGGCAGAACAACTTGAAGCAGGGAGGGAGCTTCCAGGT  
ACAGGTAGGTGAGAGACAAACAATTGCATTCTTCTGAGTGTCTGATTAGC  
CTTTCCAAAGGAGGCAATCAGATATGCATTATCACAGTGAGCAGAGGGG  
TGACTTTGAATAGAATGGGAGGCAGGTTGCCCTAAGCAGTCCCAGCTT  
GACTTTTCCCTTTAGCTTAGTGATTGGAGGCCCAAGATTTATTTTCCT  
TCTACATCACTGTGGGCAGCTGACTAGGAAAGCTTTGTAGGACTGGTGGG  
CAGTGTGAGAGCCAGTGGGGGGTGGTGGTCTGTGCCAATGGTAGCAAC  
CACCTGTGAGGCTGAGTAAACTCATTTCCCAACCTCCTCTAGCAGCCCCA  
GTGGAGATACAGATGAAGCAGACTAGCGATACAACCCAGCCTGAAGTTTT  
GTCTGGTGAGTGTAATGGAATAAAAAATGGGAAGGGTGCTGAAGAGACCAG  
CAAGAAAATGGTTGAAGAGATGGGGCACAGAAATTAAGCTGGATCAAAAA  
GGACGGAAAAGCAGAAAGGGCCGATAGAGAGAGGGGATATCTATGGGTTC  
GCGATTCTGAAAAGGACAAATCACTGGTGCTTTGAGAAGAGAGAGGGTGA  
GAAAGCAGGAAGGCTGGAGGCTGTCATCCAAGAGGCGGACATCTGTGAAC  
ATGATTCCAAGAGTCACCAGACCATGGGGGTGGCCAAAGGGAGTGCCTCT  
TCTCACCTCCTACTCTTAATTCCTGTACTCAAGATAATAAGTTCCCGA  
AGAGAAGTACCCATATTTAATTCATCTGTGTCTTCTAGCAGTACTAAAA  
ATATTATATGAAAGGTATCAAACCTTTGAGAATGTGTGCTGCTAAATTGT  
TAAGGTGCTGGAAAACCTCAAGACGTCCCTGATCTGAGCCTGAGTATGA  
GCCTGTGGTGAGCCCAATGCAGGTCTCCATTGAGACAAAGGCCTCAGGGA  
ACGGATGAGACCTAGGGACAGAGATGCATGCTGGAGCAGCATTTCCCATC  
CCTACTGCAGCTCAGGCCAGCTGACTGCTTTATGAGTAAACGTTACCAGG  
GAACACTTTGCAGTCTTAACACACATGCCCACCTGTGACCACTGATCCCT  
GTTGGGTGACCACTGACATCAGAGATTCGATGGCAGCAATGAAGACAAGG  
CTATCCTCATTAGGAAGGAAAGGAAGGAGGAGGGAGGAGGGCAAACGAAT  
CTTTCCTGCTTGTCAACCACGTCCATCTCTGTTAGGTGATTTCCCATGTG  
TGACTTTGTTTATCTTTATAATAACTCTGAGAGGTAGGTCTTGATGTCCA  
CATTTTGAACATGAGGACATCCAGCCAGGAAGTTGAGTTCTGGGGACATA  
GCTGAGAGGGCAAAGCTACATATAAACCCCTCTTTGTTTTTTCTGGCTTA  
TCCACTGAGTGCCCCCTGCAATCCACCAGCCCATTTGTGAAGTGCATACT  
ATAGGTAAGTTGGCACAGGAGGAGTGGATGTGGGCGATTTTGTACAGCT  
CTCCAGGAACTTACACACTGGTGAGGAGGGCCAGGTATGTTCTTGACCAG  
TCACAATCAAAGCAACCTCCTACTAATCAGGGAGGCTTGGTACCTGGGGA  
ATGCTATGTTGAAAGGTTCTTTTCTGGGTTTTAAATGATGGGTCTATTT  
CCTTATTCTTAAGATTGCTTTTTTTCTGGCTAGAACTTAAAGAAATTTT  
CAGTAAAATTTCCCTTCCCTGGCACAAAGTGAGCTTGAAATGAATTTCCA  
GGTGGCCTTGATACTTTAAATATTGCCTCCTATAAAATCAACCTTTAGA  
AGAAGGAAGTCAAAGAACATGCTAGATTTACAAAGGTTAATTCCTTGAA  
ATCCAGTTATCTACAGGACAATGTTGTCAAAGAAAAAATTATTTGGCCAG  
GCACGGCGGCTCATGCCTATAATCCAGCACTTTGGGAGGCTGAGGCAGG  
TGGATCACCTGAGGTGAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAA  
ACCCCATCTCTACTAAAAATACAAAAAAATTAGCCAGGTGTGGTGGTGG  
GCACCTGTAATCCCAGCTACACGGGAGGCTGAGGCAGGAGAATCGCTTGA  
ACCCGGGAGGAGGAAGTTGCAGTGAGCCAAGTTCAAGCCACTGCACCCCA  
GCCTGGGCAACAGAGCAAGACTTTGTCTCAAAAAAAAAAAAAAATTCAT  
GATATTTTTAAATTCATGGTAAGGAAGATTTCAATCAGAACCAGCACAGA  
AGATATAGGAAACACTGCAATGGGACTTTGCGGTGGGGGAGAGAGATTGA  
ACACAACCTACATATACAGCACGGGCAAGGACATATTCATAGCCAGGAAGC  
AGAGCAAAGATCAGTGGATGCGAAATTACTAAGAGGAAACATGAAAAATA  
AGGGAGCTTCTGCCTAAACCCACCTAACCGGATCCTTGCTGAAGACAGGA  
CAGGGTGATTGGACACCACTTTGGGGATGGTGGAGGATGGGGAATCCAGT  
GAGATTTCAAGGGTGATCAGATATTGAACATAGAAGGTTCTTGCTAAAAA  
AGGAGTTTACAAGAAAGTGTAACAATGTGCCTGGGAGAAGGTTGAGGAGC  
CTGACTAAAATTTGGTCAAGCAGAGAATATTGCCAAGATAATAGCTAAG  
TCTTCTGACAAACAATAGATGCTAAGCCAGCAAGGGTGATGTGCTCAGAG  
AAAGCACTGAGGGCTTATTTCTTTTCCCCCAATCTCCACTCAGTCAAGT  
CTAGTCCCCTTGTCATGTAGCCATTTGTAAGAATGCAATCAGGCAGGGT  
CCCATCTCCTAGTGACAGGACTGACTGAAGTTCTGCTGAAGAGAGTGGCC  
TGGGGCTGACACCGAGATTTGAGAGTCTGGGTTTCGCCGAGAGCTCAGT

GTAGTGCCATGCCCTCTCTCCACCTGAACGCCAGTGTGGGCAGGAACAA  
CTGCAGCTAGAAGTCTGGCATTACGCTGGGGTCTAAGACCTGCCTGATC  
TGCTAACTAGTCTTGTCCCTTGGCTATAAACTGACGTTGGCACCTGGCCA  
GAAAGATGAGCAAGAGATCTCTGACACACCTTTAAGTCCCTGTGGAGTAG  
GATTATGTTGGGGAAGGTCAATTCTCTGACTGAGCAGCAATTTGAGAAGG  
AAGTCCCATGCCGAAGTGAGAGAAGGCAGGGAATCCTGCCTAGTCAGCTA  
GAGCAAAACAGTCTGCAGGACGGGACCCAGGGATGTGATCCTCCCATCCA  
AAGGCACTGAACTAAATGACTAAAATACTTTCCAGGGCTCACGTTCTTTG  
AAGAATGGGGACTAAAATAAGACAGGAGCCAGCAAGTGAGGACTTGGAA  
GGAGATGGCTCATCTGATCAGCCTCCACTCAACAATTTAATCATCCACA  
CTGGCATGGGGACACAATATGAATAAGTTGACAGGGACCTACTCTGATTA  
AGCAGTGGGCTAGTGCAGAGACCTGTCAGTCAAGAGTGAGCAGGAGATGA  
TTTCAGACAGTGAGAACAAAATTAACAGAGTCATGTGCTAAAGGGTGGCT  
GGAACACAGAGGAGTTTAAGACTCAAGAGGTCTGGCTGGGCGCGGTGGC  
TCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGCGGATCACAA  
GGTGAGGAGATCAAGACCATCCTGGCTAACGCAGTGAAACCGEATCTCTA  
CTAAAAATACAAAATATTAGCCAGGCGTGGTGGCGGGCACCTGTAGTCCC  
AGCTACTCGGGAGGCTGAGGCAAGAGATGGCGTGAACCCGGGAGGCAGA  
GCTTGCACTGAGCCAAGATTGCGCCACTGCCCTCCAGCCTGGGCGACAGA  
GCGAGACTCCGCTCTCAAAAAAAGACTTGAGGGAGTTGTTTATT  
TTTGTCTCTTTTAAAGACAGGGTCTTTGTTGGGCGCGGTAGCTCACGCC  
TGTAAGTCCCAGCACTTTGGAAGGCTGAGGTGGAAGATCTCTTGAGCCCA  
GGAGTTTGAGGCCACTCTGGGCAACATAGCAAGACACCGTCTCTACAAAA  
AATGTGCAGGTTGAGGCTGCAGTGAGCAGAAAAACACCGTGCCTCTAG  
CCTGGATGACAGAGCGAGACCCTGTCTCGGAAAAAAGAAAAAGACA  
GGGTCTCGCTGTGTACACAGGCTGGAATGCAATGGTGCAATCATGGTTC  
ACTACAGCCTGGAACCTCTGAGCTCAAGCAATTCTCTACCTTGGCCTAC  
CAAAGTTCTAGGACTACAGGTGTGAGCCACCACACGTGGCCTCAGGAGAG  
ATCTTAATAATAAAAGGACAAATTGCCTTGATCCCTTAGGGGCAGGATT  
GACACATCAAGATCAGGCAGAAAGCCTGTGCGGAGTGGGATGAGCAAA  
GAGAAAGGCTGAGAGTTGTGAAGAGGGAGATGCAGTGCCAGCTAGGACAG  
GCCTTTTTGGGCTATGGGAGGTTTTTCAGAGGAGACCCACCTAACTAAC  
CCATAACATTGCAGTGGGGACCTGTTGAAGTCATGGACTACTACCTGAAA  
GCCAGAGAAATGGGAGGAGCCTTTCCTCTGAGGAGGGACTCTAGTCCATA  
GGTATCTTGCCACCAAATACATGGACAGGCCCTGGGGGAAGATGGTGGTA  
GCCCAGCTGGAGGAAAACCATTTGCCACCTGAACTAGCCCAGGGTAAGCC  
ACCCAGGCACTGAGGGTGACACCCATGCATGCACACAGAATCACACT  
CCTTCTTATTATTCCTCAATTCAGGGGTCTCAACACCCATTTTTTTTGT  
TTTGGGGTTTTTTTACATGTTTACATTTTATTATTATTATTATTGTGA  
CAGGGTCCCCTCTGTTGCCAGGCTGGAGCACAGTGCAGTCTGTGCAATC  
ATATTAGATTGGTGCAAAAGTAATCACGGTTTTTGTCTTAAAGTTTTG  
CCATTACTTTTAAATGATAAAACCACGATTACTTTTGACGCAACTTAAAA  
GCTCACTGCAGCCTCAAATTCCTGGTCTCAGGGAATCCTCCTGCCTCAG  
CTTCTGAATAGCTGGGACTACAGGCACATGCAATCCTACCTGGCTAATT  
TTTTAAAAATTTTTTTGTAAAGATAGAAAGTCATTTTGTGTCCAGGCT  
GGTTTCAAACCTCTGTCTTTGTGCCTCCCTCTGCCCTGTGCAAGACCTTC  
TGGATGCCCACTAATGAAGACTTCCAGGGAGAGGAAAAGTAAACATAGGT  
CCCTGATCAAGGGACCAGGGTTTATCGACCACAAACAGCATGCCAGATT  
CCACTGGCAGTCCCTAGAGGTGCGATTTGCCCAAGTGTGTGTGGAAGGCC  
TCTCCCTAGCAGTTGGTTTATACACCAGCCACAGCACAGCATATTCTCTT  
AAATTGTGAACATTTGCAAAACTCCTTGAGGACAATATCATGTCTTGT  
GTACTTTTGTGTTTCCCTTCCCCTATGTACACGCGCGCGCATGCACT  
CATGCACGCACGCGCGCGCACACACACACACACCCCTCAAACCTGAA  
TGCTTGGTGTGCTGAATGGATGAATGGCTAATGTAAGTCATTCTAAAGC  
TACTTTCTTTGGCATAACCATCACCTTTGATTTCATCTTTCTGGAACCTCT  
ATGTTCCAGATGAATTTGGAAAGCCCTCAGGAAACATTTCAAATTTGCT  
ATATGGGAGAAATGGGAGGGTCTCTCTAGAAATTTACCTGCCACAGGTAT  
TTCTGGTAAAGACACAGCAAGGTGGCACCACCCATTCTCGTTACAATGT  
CAATGCCAGTCACCTTCTGTCCCATAAACTTTATTAAAGGTGCAGAAT  
TCCCATGGAAGCAGGTGGACACCATCTGCTTCCAGCCAGCCAGGGGAGCA

AGGTGTCCACTGTGCCCTTTGTGGCAGGAACTGCGCTTCTCTACTCTCCCA  
CTTTGAGGCCTCTGGGGCTGGCCTGCTGCCTCCTCATTGACAAGGCTGCT  
TACTGAGCAGTTTCTTCTGAGCTGGACATAGTGCTTCTGGTGAGTCTCTA  
CTTCTATTTAAACCAAAGATATTCTTTCCTAAGGAAACGCTTTCCTGTCTG  
GGGGAGGTTAGCTCCAGATGGAAGTCACAAGTGATGGCATGGTAGCTCTC  
ATCCGTTTGGGTGGATGATATTACGGGAGCACCACCATGAGCCAGTCATG  
GAGGTGAACAGTATATGCCAGCCCTGAATCAGGTGCATTGACAGCAAGGG  
AGACAAGCAAAAGCTGAGGTTTGCTGAGGATGTTCAAGACTCACACA  
GCACAGAGGAGCATCCACCACCCAGCTTGGGAAAGGACTTGTTATAGAGG  
GGGTGAAGCATGAGCTGAGTCTTGAAAGACTAGAAATTAGCCAAACTACA  
AGGAGGAGAAGGAGTTTCCAGTCAGGAAGAACAGGTTATGCAAAAGCACA  
GAGACTAGAAAGAATATCACATTCAAGGAACTGCAATAGACAGGAAAGA  
TTGATGCGTGGGATAGGAGAGGAGGGCAGGGGATTCCAGGTGGGCCCTGC  
TTGCCACACTCAGGAGCTTGAACCTATCCACAAAGGAGGTGTGGAACCAG  
TAATGAATGGGTTTTGTGCAAGGGCTTCATGTCACCAGATTTGCTTTTTG  
GAGATACTTCTGTGGCTGATATGTGAGGAAGGGATGGAGGAAGTTTCCGT  
GGCAATCAGGAAAACCAATTAGCAGATGATTCAAATGGCCTAGGGGAAAA  
GGGAGGAGGACTTGGACTACCATGCAGCAGCAGAAATGGAGAGAAATAAC  
AGATCCCAGGCATCAGGAAGCGCTCAGAATGAGCCCTTCAAAGAACCTTA  
TGGTAGGTGATGGATGGATGGAGTGTGAGTCTGGGATAGCATTGCCTGG  
GAAAATACTTTCTAGTTGAGACAGGGAAGTGGGCCAGCAGAAATGGAGGG  
CTTCTTCTTTTGTCTTTAAATACTTTTATAATATTTGGAACCTTGAAAAT  
GAGCAGATATATTAGCAAAAAGCCTAAAAGGGATATTTTGAATCACTG  
CTAGTTCTAACATATAACTTTTCTAGCTTGCACACATCATCAATTAACCTTG  
ATAGCGCCTTTCTGAAACTATCATCCCAAATAGCAATCCTTGTA AAAACC  
TATTTTGAAAACCGGCCTTG TAGGATAGCCTCACAGATGTTTTGTGGTA  
GATTTTCTAACATTCTAATGTCAGGGAGTGAAAGGAATCCCGTTAGAAGT  
TGGAAAATCTCGAATCTCTATTCTGTTATTAAAGTTTTGCCGTCACAC  
AAAAGTTTACACCTTTACACAATCAGACTTCTCATTTTACATTGCTCG  
GTAATTAGAGGAAATCAGTCACCCAGAGCCTGGGTCTTAGACTTGACAAA  
ATGCACCCAACAAATCCTGAGTGGCCTTGCTGAGGACTTCTCCCAGAAGA  
TAGAAAACCTCAGTTCCAGCCAACAAGGGGGAAGCAGCTGAAGAAGTGAAA  
TTAACAAAGTCTGGAAGGAAATGACCAAATCATCTTTGATTGTGTAATA  
ACCAGAGAGTAGAATACAGCTACGACAGACATTTTGGGAGAGAAGCATT  
TATCATAGCTTTTAGAAGAGAATATTTTTCAGCATCATAAGCACACAATT  
CCAAGACAGATACTTTCAAGGGATTGTTTTGACG

>Contig53

ATGTTNNGGTTTTGGGACCCCATTCAAACTTCATGTTGAATTTTAATCTT  
CAATGTTGAGCGAGGTCTGTGGGAGGGTGATTGGATCATGGGGGTGGGT  
TCTCCCTTGCTGTTCTCAATGATAGTGAGTGAGTTCTCACAAGACCTGGT  
TATTTGAAAGTGTGTAGCACCTCTCCCCTTCATTCTCTCACTCGTCACTG  
CTCCGCCATAGTAAGATGTGTGTGTTTCCCCTTTGCCTTCCGCCATGATT  
GTAAGTTTCTGAGCCTCCAGCTATGCTTCTGTACAGCCTGTAGAAC  
TGTGAATCAGTTAGACCTCTTTTCTTATAAATTACCCAGTCTCAGGTCA  
TTCTTTATAGCAGTGTGAGAGTGGATGAATATAGTGCCATATGTTTGTAT  
TCCAGCTACCCAGGAGGCTGAGGTAAGAGGATTGCTTGAGCCTGGGAGT  
TTAAGGCTGCAGTGAGCCATGACTGTACCACTGCTCTCCAGCCTGGGTGA  
CAGCGAGACCTTGTTTCCAAAAA AAAAACCCTGTTGTA AAAATGTG  
TTCATAAAAAGTGTCTTGCTCCACACCTGTCCCTATATATCTTATTCCTC  
AGCCTCCGACAACCTACTTTATTCATTTCTTATGTATCTTCCAGAATCAAA  
AAAAAAAATCAAATACAAGCACAGTGGAATGTATTGCCCTTCTTCCCCT  
CCCTTTTGTACATCAGAGTTAGCATATCATAAATACGGTCTGCATTTTC  
TTCTTTTTCAGCTATCAGCATGTTTTGGAGAGGATTTTCATATTCGTGCAG  
ACAGCATGTATTAGTCAGTCCTTGCTTATAGGAAATACCTGAGAC  
TGCATAATTTATAAAGAAAAGAGGTTTAATTGGCTCACAGCTTCGCAGGC  
TGTTCCACAGGAAGCATGGCAGCATCTGCTTCTGGGGAGGCCTTAGGAAG  
CTTTTACTCATGAGAAGACAAAGCGGGAGTGAGTGTCTTATATGGCAGG  
AGCAGGACTGAGAGAGAGAGAGAGAGAGAGAAAGGATGCCACATACTTTT  
AAACAACCAGATCTTGTGGGAACCTGTGCACGAGAACAGCACCAAGGGA  
TAGTGCTAAACCATTCTATAAGAACTCCACCCCATGATCCAATCACCCCA

CACCAGGCCCCACCTCCAACATCGGGGATTACAATTTGACATGAGATTTG  
GGCTGGGACACAGAACCAAACAATACCAGAGTGCTTTCTCATTCTTTTCT  
ATAGCTGCCTAGTATTCTATGTCCTTTACTTCATTTAGGCAGTCTCTTGT  
TGATAGACACTTGGGTTACTTCCAATTTTTCTATTACAAATGATGTGCA  
ATGAATAATTTTGATCATTTTCCATTTACATGGGTTATGTCCATCTGTG  
GGATAAATCTCCAGGAGTGAAATTGCTGGATCAAAGGGGAAGTGCACTTG  
TGATTTTCATAGTTAGCAAATTTTGTTCTATAAGGGTCATATCAATTTAT  
AGTCCCACGCGTAATATTTAACAGTGGGGATTTCCCGACAGTTTGACCAA  
CAAGGTCTGTTGTTAAACTTTTGATTTTGTCAATCTGATGGGAAAATAC  
TAGTATCTCAAAGTGCTTTTAATTTGACTTTCTTATTACAATGTTAAGCA  
TCATTTTACTCTGCCAAGATCAAATAGTATTTTCTTTTCTGTGAACAGA  
CTGTTAAGATCCCTTGCCCTCTTGTTTTGCTGGATTTTGTCTTTTTTTT  
CAAATGTTTTGAGGCAGTTCTTTACATGTGAAACAAGTTATCTCTTTATC  
TGGGGTGTGAGTTACAACACTACTTTTCTCTGGCTTGTTTTGCGCTTTGAC  
TTTGCTTCTGGTGATTCCCGCAATTCTGAAAGTGTACTTTTTGCAATCATT  
CATTCTTATACACCCATGCTCTTGTTACGCTGGTTCCTCTAECTGAGGG  
CTTTTTCTTTTCTTTTCTATCTGGGAACATTTTTTAGAGACAGGGTCTCA  
CTCTGTCATCCACGCTGGAGTGCAATGGTGCGATCAGAGCTCACTGCAGT  
CTTGAACCTTCTGGGCTCAAGCAATCCTCCAGTGTGAGCTTCCCAAGTAGC  
TAGGACTACAGGTGCATGCCAGCATGCCCTGGCTGATTTTATTATTATT  
ATTTATTTTTTGTAGAGATGGGAGTCTCACTATGTTGCCAGGCTGGTCT  
TGAACCTCTGGGCTCAAGCGATCTTTCTGCCCCTGCCACCCAAAGTGCTG  
GGATTACAGGCGTAAGCCACCATGCCCAGCCCATGTGTGGAAATCTTCTG  
TTTATCCCTTTAGGCTTGATTCTTATGTCGTTCTCCTCCCTCCTTCTG  
CTACTCCTCTTGTTCTTTATCTTACTCTACTTGTGATGTTACCTTGTTTC  
TGCTTATAACTAGCTGCCTCTCCTATCTGAGGAGGGACTTGTGACTGTTT  
TCATCTCTGTACTCCAGGTCCTAGTACATAGCGCTTGCTCAACAGATGT  
TTGGTGCAATGTAGATAAATCAATGGTAGCTGTTAATACCAGTCCTGAC  
TCCCTGCAGTGCTTCCAGCTGATCCTGTTCCAGATGTGCACTGAATATCTT  
TCTGTTGAACAACAGAAATAAAGGGGATGGGTGAGGAGGATAGTCTTCGG  
TGGCCAAGGATATTTGTAGGTACTTTGCAGCACTCAGCAATGAGGAGTGG  
GCTTTAGTCCCCCAAGAACTCTCACAGCCCTGTTTGTCTTTACTGTTTCTG  
TGTCAAATCCAAGACAAGTCAATGATCAGGAAAGACCTTTTTTTTTCTTC  
AGTGAAGTTTATTTTCAAGAACATTGAACAGTATGATATTTGCTCATTAT  
AAATATTTCCCATTTAAATAATCTGAGCTTATATATTTTTCAGTCTTAATTA  
AAGGACTTGATTTAAAGAGAGCACACCAGTCCAAATTGAATTGATTCCAT  
AGCTATTTAAAACTAGGCTCTTTTACAGACACTGCTACTTCTTGCCCCCT  
TTGAATAAATTAGACCAATGAATAAAACAAACAAATAAATAAATAA  
ATAGGGAAGCGGTTGCTCATCAGAATGTGGGAGCGAATGACAGAGGGTTT  
CTTAGAACCAATGTGGCCGTGGTTTCTGTGAGGCGGGCTTTAAGTGAGT  
AGGAGAGGTGAGAGAGGCTGGCTCAACAAAAGGGCTGGGGATTGGCCCT  
GAAAGGAGAGAGCTGACTGTCTGGCTGATGGACAGGAGATCCTCTTAGC  
ACTACCCTAAGGCAGGCAGTTGGGCATTGGTGTAGACAACAGGAAAGTCC  
AGGCTATAGCCGTACTCAAAAACCTTTCTGTTCCCTTTCTGCCAGCCCTA  
GGGATTGAGTCCACATTGAGCAGGACTCTCTGGGTACAGCTCTCTTTA  
GGAAGACACAAATTGCATGGTGAAGTCAGTTATATCCTGGCCGCCTTTGG  
TCCCTCCCAGGAAGACGGGCATGTTTTCTGCTTGAGAGGTGCTGATGTAC  
CAGTTGGGGAACCTGGGCAGACTCAAATTCAGCTTGTTATTGATTTCTAT  
CTTGTTGAAGACAAATCGCTTTTCCATCTTCTTTGGGTAATTTTTGG  
GATCTACACTCTGCAGCGAAAGAGAAAGAAGATTTTTGTGGGGCAAGGG  
ACAAAAATGCTATGGGAAAGATGTTCTTTGGGTTGGCCAGAAAGGAACT  
GACGAGCAGGTACATGATCAGGAGCCACACTCCTGAGTTGTAAGTGGGC  
CCCCAACTTTCTGTGTGATTATTAAGAGCCCTTCTTCTTTTCTAAAC  
TTAGTGCCAAATGCTGAGGAGCATAATGTAGGTGAGAATTTTTTTTTTT  
GGGGGGGTGAAATTAAGCTAGAGCTTCTTGAAGTACCTAGTTTCCAGGG  
GCTTTTTATTGTATTTTCTTATGGTCCTAGAATGACATCAACTTGGA  
ATGAAGCTTTTGTGAGAAAGCTGGAGGTGATAGTGGTGGTGATTTTGGG  
AGTGGAGTGGACGTGATAATGGGACCCTTTAAGTCATCTATTTCCCAAGG  
TGTCTATCAAATGAGAGCAGCCCTAACAATATATAATCTGTTGGGGTGT  
AACTATGGTAGGACATAATAACATCGGCAAAATGATTTAATTTTCTGCAG

CAGGATTGAAGGTTGCAAGCAGTTAAAAATTATGTTAAATTTATTTACAT  
TAATGCAAAATTGTCAAATAGACCTGTTCCAGCTTTTCCTAGGGATGGG  
GGCGGGGAGAAGGTGGTTGTCTGGGAATAAGTGGTAGCAGGAGGCTGAGA  
AGGGCTTCATTCCATAGCATTCACTTACCTCCAGCTGTAGAGTGGGCTTA  
TCATCTTTCAACACGCAGGACAGGTACAGATTCTTTTCCTTGAGGCCAA  
GGCCACAGGTATTTTGTCACTTTCTTCTCCTTGTAACAAAGGACATGG  
AGAACACCACTGAAGAAAGAAGGGGGTCTTGTGGTTAGGGACACAGCAGT  
GCAGGGTCACCCCAACCCCTAGGCCCATGAGTAGGATACATGTAATTG  
GTAGCCTCTGTGGGAACCCACAGTGAGGTTCTTGCCCTAAGACACAGGA  
TAACTTGACTTCTCACAGACAATAGCAGGGTCATTTGTTGATTTAGGGT  
TTCCCTCAAAGGCCTGAGGGTTTCTCAGAGCCTCATAGCAGTAGGAACG  
GAGAATGAAAGAGGGTCTACATTTTAAATGCTGAAGGAAGGAAGGAAGGA  
AGCCATTGTGTCACTGGCTGGCAATGTGCCCATCCACAGGAGCGGAACAA  
CTTGATCAATGTGGAAGGAAAGGAAAGAGGTGAGGCTGTACTTCTGCCAG  
AAATCAGGCACCAGAACTGTTTCAGGAACAGAGAGTAGCCCATGGGAAGA  
AACTGGGAGAGGAGAGGCTGAGCTGGGAAAGTGGCTCCAAAGAGAGACAC  
TCATTTTGATCTTCTCAGTCACAGCAGTGTCAATTGGAAGGCCCTGGGA  
TCACCTTACTACCCGATTCCAAAGAAACAGGATTTTCTTGGCCTGGCTG  
AGAGCAAAATAGCTTCCCCCTTGAGTGAGGCTGTCTTCAAAGTCAGCAGC  
CTTAGTTGCCCACTCCTGTGCAGAGGCTTTGGCTACTGTGGCACGATG  
CCAGGCAGATCACCACAGCTAATGATGGGTTACCCGCACTTGAAACTTTT  
GCCCCGTACAGCGGAGAGATATAAGTTCTGCTGGGCGGTAAAATTTCCC  
TACAAGGAACCACCTGGCATTGGGTGGGACGATGTTGGGGCAAGGGGGG  
AAGACTGGGGAGGGGGATGGACACATTATCGCTCCAGCACTCTGTTTCA  
GCCTCAACAACAGGAAGAGAGAACCCACAGGCAGTTAGGCCATGTCCATC  
AAATGACCCCATATTGTGGAAGAATTGACATTGCACTATGCCCAAGAGAC  
TTGGGTGACATGCTCCTGGGAGTGCTTGAGCCGTCTAATTTCTCAGGGT  
CACACTCCTGTTAACAATAATGCACTGGCCAGTGCAATCAAATGTGCCATTT  
CTAGGACCAAAGTTTGTATATTCTTTTTTAATATTTTTTTTCACTTGTGT  
TGATCATTTGCCTTAAATTAACCTTTCTACTTTGTTTAAACATGGAGAAT  
TAGCAAGCTGCCAGGAAGCCAGGCAGGGAAACCAGGATGTTTCCATTTAC  
CTTGTGTGCTCCATATCCTGTCCCTGGAGGTGGAGAGCTTTCAGTTCATAT  
GGACCAGACATCACCAAGCTTTTTTGTGCTGTGAGTCCCGAGCGTGCAGTT  
CAGTGATCGTACAGGTGCATCGTGACATAAGCCTCGTTATCCCATGTGT  
CGAAGAAGATAGGTTCTGAAATGTGGAGCACATGTTGTTTAGGTATAAAA  
TCAGAAGGGCAGGCCTCGTGAGGCAAGGTGGCAAAATTTGATTTCTTGGA  
GGACACTGAGCATATACGGTCAAAGTCTGATGACAACACCAGTAGGGAT  
GAAGCTGGGAGTGGGGTGGCTAAGAACACTGGACCTGACACTATTAGACA  
TGGGTTCCAGCTTCAGGTCTATTACTGCTCACTGTGGCCGAGCAACAGAG  
CTACTTAGGTAAAATGGTGATGGTCATAACACTAGCCACAGGGAGGTTA  
CGAACCTCTGGTGACAATGTAAGTGAAAGGCCCTGAGAAAGAGTGAGGG  
AGTTGCAAATGTCACTAGCCATCAAGATCTTCTTTAAGAATAGTTTCCAC  
TAAAGAGATGATTGCTTTGGTTTCCAGCCTTCTTTGTTTGTCTCCCCGC  
TGGGCCTTCTACCTTTAAAGGGCTTTGGCTCTGGGGGAATTGAGTTGGCT  
GGGGCTTGATGACTTCCAAGAGGACACAAGTGGAGATCTACTGCCTGCTC  
TTGGCTAACTACCTTCTTCAAAGATGAAGGGAAAGAAGGTGCTCAGGTCA  
TTCTCCTGGAAGGTCTGTGGGCAGGGAACCAGCATCTTCTCAGCTTGTC  
CATGGCCACAACAACCTGACGCGGCCTGCCTGAAGCCCTTGCTGTAGTGGT  
GGTCGGAGATTCGTAGCTGGATGCCGCCATCCAGAGGGCAGAGGTCCAGG  
TCCTGGAAGGAGCACTGCGGAGAGAGCGAGGGAGGGAGCCTGGTGAGGTG  
GTCCTGCCAGGAACCATGCTTTGACATCAGAGAGTAGAAAGCTCAGAGAG  
GAGGAAAGGGCTTGAAAGAATCCCGAGCTTCTAAAGATCATCCCTCTCTG  
GGCCAGGCGTGGTGGCTCATGCCTGTAATCCCGACACTTTGGGAAGCCGA  
GGTGGATGAATCATTTAGGTCAAGACTTCAAACCAGCCTGGCCAACATG  
GCCAAACCCCTTCTCTACTAAAAATACAAAATTAGCTGGGTGTGGTGGG  
GTGACCTGTAATCCTAGCTATTGAGGAGACTGAGGAAGGAGAATCGCTT  
GAACTCAGGAGGTGGAGGATGCAGTAAGCCAAGATTGTACCACTGCACTC  
CAGCCTGGGCAACAGAGTGAGACTCTGTCTCATAAAACAAAACAAAACAA  
AACAAAACAAAATAAAATAAAATAAAATAAAAGATTATCCCTCTCTGAA  
GCTCAAGGAGGTTAAGGGTGACTCAAGGGCACACAGCAGGTTAGAGGCA

FIG. 4 (48 of 61)

702/118

GACTCAAGACTAGAATG<sup>1</sup>GGGCTTTCTGACACCTTACAGGCTATTCTTTT  
AGAATAAATCCCATTCTACTTTGTTTCATCTTTTGTACATGCCCCACC  
TACACCATACATGTATACCTTCTCTATATCTTTTGTATCCCTAATGCTG  
TCACACTATGATTTGCTTTTTCATGCAGATGACCATAACATTTTCCATTC  
ACCTATGCTCACTCAGCAAGTATTCAATTTTCTACACTGTTCTTTT  
TCCTTTTTCATAACACTGTCTCATAGGCATTCTGCAAATCCTGTGAGAGT  
ACTTTTGTGAAATGTTACCCTTTCTCTTATTTCAGAGAAGCTCCGTAT  
TAAGGCTTCACTGAGGTTGCCTTAAGGCATGATAATGGTTCAAAGGCTTG  
AAAGACAGTTAAAGAGACCTGTAAGTGCACAAAAGAAAGTTGAGCAGGAG  
AGAATTTCTTGCTGGAGCAGAGCCAAGCTACTGGAAGAGGCAATGGGGG  
CAAAGGCCAGGCAGACAAGCCAATGGGCTCTCCACAGCTGCAGCCAAC  
AAGTTATGCCAGTCTTAAACTTCTAAAGAAATATGTTTTTAACAAGATT  
GAGGACTGGATTATGAGGCTAGGGGAGGCTATCACAACTGGAATAAAAT  
AAAGCCAGAGAAAAGTGGCTGCCTTCCAACCTGCACAACTGACCTAGCTA  
GGCTGATGGCTGGGCCACCTAGGAAGGCTACTGAGCATCATATAAAACAG  
AAGGGACAGCAGGAATATAACATGGCTCTTGTAAAGGATGAGTCTGAAAA  
ATGACCATTTGCTGCCCAAATGCCCTTAGCTACAACCTGAAAATATTTT  
AACTGGAGGTTGCAGGATGCTGGAATCTCAGAGATCATCCAGCTCAGCCC  
TTTATTTTTCAGATGAGGTCCAAAGCGGGTAAATGACTTGTCAAGGTCA  
AACAGCAAGTGAATGGTTTCTTTCAAGTCTCAATTCATCTTTTGT  
TATCATCTATGTCTTGTGTTTATAAGCTTCACCCAGGTAGCAAAAAC  
ATTCTACTCAAAAGGGGTAGACATATGTTAGTTCTCAAGATCATCTCTG  
GTTTCAGAGTTTAACTCAAGTGATTGGCATAGGCTGAATCCATCTCTTAA  
AAGGATAATCAAAATTTATGTTGAAGACTTGGTTGTCTTCTACTATGAAA  
TGGGAAACATTATCACTACTCCTCCCTGTCAACCAAGTGTGGCCACC  
ACCACCAACGTTAGTGAGTGACTGTGGTGATATGATGACCAAGTGGCCAG  
GTCAGCAAGTGGTGCAGCCTGTGTCTCACTGGAAGAGGTTAAAGTCTTTC  
TAAAACAAAATACCATGGCATCAAAGTGGCCAGAACTCCCTTCTTTGAG  
CTTTCCTGTGTTAGAGCCCTTCTTGGGTTGGGAGTTAAACCCATAGTC  
TTACTTTCATGTTTATAGGGCCATCAGCTTCAAAGAACAAGTCATCTCA  
TTGCCACTGTAAATAAAACAGGGACATGTCTCAATTATGTCTTCTAAACA  
GGTTTATTTTCTTCCCTGTGTACAAGACTTGACTGTTTATAAGAACT  
GCAAACAGCCTGCCTCTCAAAGCTGCCTGAAACACCTGGCAAGTTTCACA  
GTGATATGCGCAGAACAGTCCAGAAGGCAGATTCTAGGCCTGGCAGGTGG  
GCACCCTGGGTGCTCCCTGTTGGATCTTGAGGCCTAACCTCTAGCCAGC  
AGAGTCAGCTAAAATCTGAGCTCTCCCTCTCCCTCCAAGCCACACTTTGC  
AAAGGGATTCTTGTATTGTGGGCTTGAATCTTTCTCCCATTTGCCT  
CTGCAGGAAGCCCTTGCAACAACACATCTGGATAGCCTCCAGGTCCCAAG  
GCTGGAGGGACTTGTAATGGGAAAGTAGTCTTTAAATCAGATTTACTTGG  
CACCCTGTTGCCACTGAAAGAGGGCAATTTAGGGGAAAAATCTGGTCTCC  
AAGCACAGATAACACTCTACTCTTGAAAGAGGAGACCTGCTCATGTTACT  
GGTCTCAGCGTCTCCACTGACCTGTAATAAGCCATCATTTCACTGGCGAG  
CTCAGGTACTTCTGCCATGGCTGCTTCAACACCTGTGTAAAAAGGAGAA  
AATGAGTGACTTCCCATGACGGCTACGTTTATGTGTGATTTCTCTCAGC  
ATCCAGTGATGGCAGTCATGCAAAGAAATGATCTCTGAGTAAATGAATG  
AATGTGTGAAAGAGAAGTCTTTGGGTCTAGAGAAAAGCATTTGCTAAAC  
CAAACCCCACTAGCAATGTATTGGCTAGGAGAGCTGGAGCAGAGGCTTT  
GACACTAACCTTTAGGGTGTGAGCTGTTAGATAAGCAGTATCCATTCCCA  
GAATATTTCCCGAGTCATAAGCATTATATTACACCTGGCATTTTTCAAA  
AAGCTGAGAGAGGGAGGCAGAGAGGGAAGGAGAGGGAGAGACAGAGAAAG  
AAAGAGAGAGAGAGAGAGAATATGCATACACACAAAGAGGCAGAGAGACA  
GAGAGACTCCCTTAGCACCTAGTTGTAAGGAAGATTAAAGTCATACTTGA  
GCAATGAAGATTGGCTGAAGAGAATCCAGAGCAGCCTGTTGTGCCTTGT  
GCCTCGAAGAGGTTTGGTATCTGCCAGTTTCTCCCTCGCTGTTTTTATAG  
CTTTCAAAAGCAGAAGTAGGAGGCTGAGAAATTTCTCTGTTGAATACCTG  
ATTTACAAATCAAGTTAAAGGAAAGGGGAAAAGAGTATTGGTGGAAGCTT  
CTTAGGGGAGGGGACTAATAAAGCTGAGATAATTCTCTGGTTTATGGAAGG  
GCAAGGAGTAGCAAACTATGACACATTTTGCAAATGTATCACCATGCAAA  
TATGCATTGTTTTCTGACAATCGTTGTGCAGTTGATGTCCACATTAAAA  
TACTGGATTTTCCACGTTAGAAGAATGTTTAAATTTAGTATATGTGGGA

FIG. 4 (49 of 61)

103/118

CAAAGTGGGAAGACACACAGATTTATACAGCACATACTTTTCTTCATTCA  
CTTCTTTGTACTTAAGTTTAGGAATCTTCCCCTTACAGATGGATAAATG  
GGTACAATGAAGGGCCAATAGCCCTCCCTGTCTGTATTGAGGGTGTGGGT  
CTCTACCTTGGGTGCTGTTCTCTGCCTCGGGAGCTCTCTGTCAATTGCAG  
GAGCCTCTGAGGAGAAAATTGACCTTTCTTGGCTGGGGCAGAGAACATAC  
GGTATGCAGGGTTCAGGCTCCTGACGGAGTTGGGGCAACCCTGGAGATAA  
GCTCACACAACCCTGCAAGACCAGGTGCTGTTACCCTAGCCAATCTCATG  
GATGAACCAGATCAATGCCAGATGAGCTCTGCCTAAAAATGATTTTTTGGT  
GAACTCTGAAAAGTGGAAATATTGTTTCTGTAAGAATATCCATCTGAGACT  
CTATCTCTTGGTAATACCAAGAGTTATCAGTTTCTCTTTAACCAGAGACAC  
CAGCAAAGTGCTGCTCCAGGGTACTGCCAGGGGAGCCCTCCATTGTA  
GAATGAATGAGAGTCCAGGTTATGAACAGTGCCTGGAGTGTAGGAACACC  
CTCCTTTGCCCTCTTTGACAGGTCTGCATCATAACACTTTTTTTTTTTTTT  
TGAGACAGAGTCTCACTCTGTGCGCCAGGCTGGAGTGCAGTGGCACGATC  
TCGGCCCCCTGCAAGTTCCGCCTCCCGGGTTCACACCATTCTCCTGCCTC  
AGCCTCCCCAGCAGCTGGGACTACAGGCACCTGCCGCCACGGCCGGCTAA  
TTTTTTGTATTTTTAGTAGAGACAGGGTTTTCAACATGTTAGCCAGGATGG  
TCTCGATCTCCTGACCTTGTGATCTGCGCCGCTCGGCCTCCCAAAGTGTT  
GGGATTACAGGCGTGAGCCACCGTGTCCAGCCTGTAACTTCTTATAGC  
ACTGAGTTGAAACCTTGCTCCTCCTGGTTCTCCTCAGGAACTGAAATCTT  
TTTGAGCCAAAGTCTAGCAGAGTGCCTGGCATGTACATTAGGTGGTAGAG  
TTTGCTGCTTGAATGGGTGAATGGGAATTTGACAGCATTTTTTATTCAAAT  
TAGTATGTGCCAGGTATCGTGCTCGCTCTGCATTATCCAAGGGAGTGAGC  
CTCTGTGCAAGTATTTGAGACACGAGGGAAATAGGTTCTACTGTGGGAAA  
AAGAGCATTTCATGGACTTGCTCTCCAAGCAGCCTTCTGATTTTTAATT  
GGCTCCCAGTATCTTGATATCAGGAGTCAGTCACAAGAACTCCATCTTTA  
GTAAGTTATATTTCCACAGGAAATCTAAAAGCTGTTCAACATGTTAGTT  
TCCTGTGAATTTGATAAGCCATAATCCATTCTAACACTGAGCCCTCCTG  
AAATTTGGTGTCTGGTCTGTCAGATAGCTAAAAGCCCTGTCTGGGTGGCC  
TAGGGGACTCCTCTGTTTTCCTCCACAGGATCCACTTTGCAAATTAACC  
ACTGGTTCTCCCGTTGTAGGAACTGCCACCTTCCTCAGAGCCTGTCTTTC  
TTCCTTCCTTCCTTCCTCCTCTTTCTTTTCTTCTCTCTCTCTTTCTT  
TCTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT  
TCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT  
TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT  
TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT  
TTTGTCTCTCCCTCCCTTCTCTCTCTCTTTCTTTCTCTCTCTCTCTCTCT  
CCTAGACAGGATCTACCTTTATCCCCAGGCTGGAGTGCAGTGGTACAAT  
CATGCATTCAATGCATGATCACAGCAGCCTCAAACCCTTCCTCAGAGTCT  
TTATGCGGCAACCAGCAGGGTCTGGAGGGTTGGTGGCTCTGTGAAGTCTC  
CTGACAGAACACAGAGATGTCTTTGGTCTGTTGATGTGATTACAAGCTGA  
ACGAAGGAGGATCAAAGCCAGTGACAGGAAGGGAGATATGCAAGGGACCC  
GAGCATCAGCTCTGAGTTAGTCCATTCTGCTTCTGGGACTTGGGATACAG  
GTCAGAAACCTTGAGCTTCTACTTCTCCATCTTCCAATTGTAGCATCCAG  
GACCTCAGAATCTGCCAGCTAAGAGGAGCCGTAATGATTGTCTGGTGGGA  
TATGGTGGGACCACAGAGATGAAGACATGAATAGCTATTTGAATGTGAAC  
AGCAGACGAAGAAATCAAGGCTAGGAGGGTGAAGTGAAGTCAATCAATAG  
CACAGTGTGGTTGAAGCAGCACTAGTATCCAGGTTGCATGAGCCCTGAT  
GCTTTTCGCTCGAGGGAAATTTTGGAGCCATGGGGCAATGCCCCCTGACGT  
AACAGTCTCCACAGTCTGCCATGTCTCATCCTGGCCCTGTAACCTGGAC  
CCAAATCTGCTACCATCCCATCCATCTCAGGAAGTGAAACCTCTTATGTC  
AAATAGGTTGTGCAACGTATGTATCAGATCCTGTCTTCCAAGGAGACCG  
CTCAGGCCACAGCACTTCTTCCGATCCCCAATGAGCAGAAAATATCTCG  
CTATAAACATAGTTGGCACTAAGGGAGGGAGTGGAAGAGTGATGATGATG  
TAGATGGTGATGTAGCCCCAAGGAAGTGAACAAGCAGAGATGGGGAGCT  
GGAAATGCCAGGATGCTCCAGCTTTTGGGGGAATTATTAGCTCTTGAGTC  
ACTAAAGCCTTTCTCAGCTGCAAGTTCCTCTTTACCCTGTCAGGTCATT  
TTCCAAGCAGGAGACTGACATTTATTCAAAGCAGCAAGTGCCCTGATAC  
CATCTTGTGTGTAATCATGGGCTTCGAGCCAGTTATCAAGGTTGATCTC  
ATCTCATTGGTCTTCAATCATTTTGAACAAGAAGACAAGCAAAATAATCA

TGGGTTAGTTCTTATATTATTGTGTGTACATGCAGTGATGTCTGTTCTTT  
GTAGTGAGCTGTTCCCTTCCTTGTTCAACCCTCTTGCTTAGAACAGAACTAA  
GCAATCTGCCCAACATTTTCCCAATTTCCCATCTCATTCTTGGCACT  
GGCTTCCTAATATTTGTTCTTATGAGTCATTTTCTTGATCATTTCATG  
AGTCCCTCTGGGATCTTAAAGTATGAAAAATGTTGTGTGTACCCACACCT  
GTCTTTGTGGATATTTCTCTCCTTTCCCTTCTGCTTCTGGGATTATTTGG  
GAATGGGCACTATGATTTTTATCATATCGCTTCCACTTCCTTTATGGCAT  
CATCTCCAATGGGCTTCTTCTCCCTCTTGATCCAGGTTCTCAGATTGGG  
GACATGCAGAGTCCAAGGAACATTCCATTCTCCTCCCTGGTCTAGAACAA  
GGAGGGCTTAGATATATGAGCAGGTGGCTGGGGCTGGCGAGCTATGTAGT  
CTCCAATGGCTTTTCCCTGATGTGCGAGTTGTTATGTCACTTCTGGGAGA  
CCAATAAGACCTTGTCTTCTTGGATCCATCAGAAAAAGCCCCTGGGT  
GGGTAAGATGGATGGCAGGGCTCTCCTACTCTATGTCTTTTCTCACACCT  
AGTGGGTATAAGAGAGGGGACCACAAACAGAGGGGGCTCTGGTACCACTT  
ATCCAGGGTCTGGAAACATTTTCTGTAAAGGGCCAGATAATAAATGTTT  
AGGTACAACACTCAACCTTGCATCATTTCAGAAAAGCAGTCAGATAATA  
CATAAATGAATGGGTGTGGCTGGACTTGTCTGCGGTCCCCTGTCTTATA  
TCATTGTATTATATCATTTTTTCTTACATACAAATTTAGAAGCAATACTT  
AAAAAAAAAAGCCGTCCTTTATTGAGCACCTACTAAGTGCCAGGTACCT  
TTTTTCCCTCATTATCTTATTAACCTCTTCATAATAACCTTTAAAGTAGA  
TAATATTGAACCACTTTGACCTATGCAGAACTGAGGTTGAGACAATAAAT  
TATTTAAGACCGCACAAACAGTAAATGCTGGAACCTACGACTCAAATATGG  
GTTAACTGAACCAAAACCAGATCTTTATTTCTCACTTTTAATTGTTACAT  
ATGTTTATTGCCTCATCTCCTGTCCACATGGTGCCCATCGGCAGACTCCT  
TTCTCATTCTCAGTGATTGAGTGACATTCTAAACTACATTGGCCTGGCAG  
ATTCACCTCTGTCCCCTAAATGTTTCCACATTGTCTTTTAGGATTGAGA  
TCCTCTCTGTTCCCTTGTCTTCCCTCCTTTCTTCTCTGCGGTGACGTG  
CTGTGTGAATTTGTTTCTTTCTCCTCTCAGGGTAGTACTGGGACTTTCCA  
AATCAGGGTTTTTAATGATCTCTCTCNCTTTTCTGAATTTCTTCTTAT  
TCCCATTCAGTTTCTCATCTATAAGTGCCANCTTGTGTGCTGGAAGATAT  
CCCTTGTGCGGGATTNCTCTTTAANAATTTGTCNNNACC

>Contig54

GTGATCGTCAACCTCCCACCCTGTAGGGCCTCAAGCATTGAGGACAATCA  
CTGGCTGCCCATTAACCCAGAAATGTTGCCGAGACAGGAGGCCGTGGCCC  
AAGTTCTGGAATGGGGTATTATTATGTCAGCACAAAGGCCTTTGCACAA  
ATGAAGGCTTTAAAAATGCAGTCCTAGTCAGGTGGAGGAGGGCTTATAGG  
ATTCCCAGGAATCTGGATCATTCTCTTGAGAGCTTTCCTTGTCTCTGTT  
AAAACCTCACATCGTACGGCCCAATAACAACAAAAATGGATGTAAATTC  
TTGAAATAACTTGTGGATGGGGGAACAAGGCCACCCCCCAGATCTGCCA  
AAGCTTCAGGTGAGGGTCCCAAATGCCAAAAAGTCTGGTATCAGAGAGG  
ATGGCCAGTGACNTGGGGACACATGCCCTTTGCTGTGTCACTCAAGGAGC  
AGCAGCTTCGGCCCCGCACAGTGACCAGGACCCTGGCTTCCCACGCTGGG  
CAGGAGCTGGTGTCTGATGAAGGGAATGCCTGGCAGCACGTGCTGTCTGT  
CTCCTCGTGTGACCTTACCTGGCTTTGCTGCGAAGAGGCCACTTGCAATTT  
CTTTATTTTTTATATTTTTTTAATTTTTTAAATTTTTTATTTTATTTTA  
TTTTTATTTATTTATTTATTTTTTAATTTTTTTTAAATTTTTTAAATTATG  
CTTTAAGTTTTAGGGTACATGTGCACATTGTGCAGGTAGTTACATACGC  
ATACATGCGCCATGCTGGTGCGCTGCACCCACTAATCGTCATCTAGCAT  
TAGGTATATCTCCAGGTTAATCCCTCCCCCTCCCCCACCACCAAC  
AGTCCCCAGAATGTGATGTTCCCTTCTGTGTCCATGTGATCTCATTGA  
ATTTCTTTAAAGGTGGAATCTCTCAGTGGGGTCTAATCTGTTCAAGAAATA  
TCAAAAGAGTATCCTTGGAATGACTGGAATTCAGAGTCATCTGGTAAT  
CCTCATAAAACAACCTGGATGTCTCTCAGCACATCTCCACCTTGAAC  
GCAGGAGGCTGGTTCAAATGGAGGAGCATCGCTCTACTGCACTTTTTTTT  
TTTTTTGGCCTAAAGTGCAAAGGGGATACGTTTCATGTAAATAAATCAA  
CTGCAAATCGCTAGTTATGCTGAGCCCTGTCCCGTGTGTGGACACAAAG  
GAACCAAAGGCTTTTCTCCCCGCCCAACACACACATAACACACACACAAA  
ATCATAAAACATACATACCCCCAACACATAACAACACACACACACACA  
CAAAATATATACACACACACACCAACATGCCACAAACCTGTGTCC  
AAAAATAAATCCTACTGGTGGGTTTGTGGTCTCCCTAACTTCAAAATGA

AGCCGTGGACCTTCGCAGTGAGTGTTACAGCTCTTAAAGATGGCATGGAT  
CCAAAGAGTGAGCAGTAGCAACGTTTACTGTGAAGAGCAAAGGACAAAG  
CTTCCACAACCCAGAAGGGGACCCAGCAGGGTTGCTGGTTGGGGTGGCC  
AGCTTTTACTTCCTTTTGGCCCCCTCCCATGTTCTGTTTCCATCCTATCAG  
AGTGCCCTTTTTTCAATCCTCCTGTGATTGGCTACTTTTAGAATCCTGC  
TGATTGGTGCATTTTACAGAGTGCTGATTGGTGCCTTTTACAATCCCCTT  
GTAAGACAGAAAAGTTCCTGATTGGTGTGTTTACAATCCTCTTGTAAGA  
CAGAAAAGTTCCTCAAGTCCCCACTGGACCCAGGAAGTCCACCTGGCCTC  
ACCTTTCAACTCCATAATGGCATGAAAATACATATGTTGTACAAAACATA  
CATACACAAAGTATACATGCATCTCCCCAAATATACACATACCACAGAAA  
CATACACACAGGAAGTACAGTACCTGTCAAAAGTCTGCATGGTGAATTGCC  
TCTGCAGTGAGTAGTTAGAAAAGTGAATTTGTTTTTCAATAAATTGGAGT  
CCTTAAAAATCGTTGTAAGATAGAAAATTTTAAAAGTATATAAAATAAA  
ATATGTATGTCCTTTGGTCTAGCATTACACATGTAGGAATTTATCCTAG  
TGGAGTAATCAATGATATATGCAAAGATTTGGACAAGCATATTAAGCACA  
GAATTATGTATGCATATGTGTGTGTATATATATATATCTATACATAT  
AATAATGTAAAAGTGAAAATAACTCAGATGTTCAAAATTGAGGATTAGTT  
AGACTATGTTCTGTCATATGTGACATACAAGTTAGCTGCCCCCTATTCT  
CTCGAGCTTCAACCTCCTATAAACAGTGTCCCTTGTATATCAGTATTGGT  
ACAGATAATCGAACTTATTGAGGTTTTACATGGGGCAATAAAGGCAAGAG  
TTTATGAATACTCCATACTACACTAGGTAGCACCCTTATTAAAGACAAA  
CTCTTCTCTCTCATTTCCTTCCCTTCCGGAACCACTTGGTTGAATCTCT  
ACAAGTCTCTATTGCAACTGCCTCAACATGGCACCTCCTGCATCTCCA  
TCTTCCCTGTCTGAGAGCAATGGCCTGCTGCCCCACACTCACATCCTC  
ATTCATTCCAGAAGTGAGCACCACAGAAGTGCCTACAGTTACCCCAACCA  
CCTTCTTAGAAGATAAGTTAGTGTGTTTTGACTTTTTAAATTTTTAC  
TTCTCTTTTCTTCCATCTCATCCCATCCCAAGAGGTTTATCAAGAA  
GTTCTCTAAAGATATGTGTCTCCTTATGGAATTTAACAGAAATCAGGGAT  
TTGTATTCTAGCCATCAAGGGAATAACATTTTTCCAGGTCTTTAGACAAA  
TAATGGAATACCTTGCAGTAATTAGATACACTATTGTAGAAAAGTATTGA  
TGAAATGGAACGATGTTTGAGATATCATATTGAGTAGAAAAGGCAAGATA  
CATTAAAGTAGGAAATGTATCTTACAAAATAATTTGTCAGACACACTCCTA  
TATTTGTATGTTATATAAATGCGTATGTGAAGAAAGGCTAGAGGATGAGA  
CCACAGTCTTCGGTGAAGTTAAGAGATGAGGCTGCAGCATGCTCAGAAA  
GGCCTGGGTTATAGTTCTTCCAGTAATTAAGGATGTGATCTTGGGTAAAT  
TGTCCATCCTCTCTAAACTGCACCACCTTTTGTCTGTAAAACAGGAAGGA  
TGGTATTACCCCCAGGGTCATCAAAGGATTTGGTTGGAGAAAAATAAAT  
AAATGGGCTTGAGCCAGACCTGGCACAGTGAGAGCACAGTGGTTGACTAT  
TGTGCTGGCCTGTTGTTCTGTGTTATTGACATGCTGCTGGTGGTGGTCC  
AGAAGCTATTACCTTAATTGGTTATGTGGATTTCCCCTCATCTGAGCAG  
CTGTGTGTGGTGTGTTGTAACATAGCCATACACAGTAAGTACAAGGGCA  
AATGTGATGGAAAAATGCAAGGAAGTGCAGATAAATAGCTAATGGGCTGT  
AGAAGGAAGCTAGTCCTTGGAGGGCTTGATCAAGGAAGGTCCTTTTGCAT  
GTCACCTTTGAAGAAGAGGGGACATAGAAGAGGTATAGTGCATCCCGGAG  
TGTACCTGGAAGGGAACATGAAAAGAGGACATTTTTCTCTGGGACATGGG  
GACTCCACTTGCATGAACTCTGGAATTGGGGCAAAGAACCATCATGAGAA  
CAAGGGCTTCCTTGAACCTCCAGGCTCATTGGCTGATCTAAACCCTGTG  
TCCCCCTCTTCTTCACTCTCCTCTGTTTTCTATACCTGTATTATTGGAC  
TGGACTGGAAGCCACCTGATCTATCACAAGTACCTTGAATGTGTTGAAT  
AGGTGTGGCACAGTCCTTAGCAGAGTGGCACTACCCCCACAGGAATTTGT  
TTATACCTTTGGCATGGAATAAGCAGGAAATGAGTGATCACTGATAACT  
GAGGATGCTATTTATTATTGGCCAAAGGAATACTTGTGTTGTATTGTCAT  
AACCCTCACAACTGTTGATTACAAATGAGTACCAGACCTAGCTCCTTC  
AAGTAAAGGATCCTGAGAACTGAAGGCAAACAGAGCTCCAGGAGTCCAAG  
ACAGAGCCACAGACCAGAGGATCCCTGGCCAGGTAGGTGGTCTCCTG  
CACTGGCTTTCAAGGCCAACAGGATGGATGGGGAAGTAGAGTAGCATCTG  
CCCATCTAGACCTTGCTTTTTTATCCCCACTGGAAGCACATCTGAATTC  
TAAATATGATCTCTGAGACCTGCCCAGAACCTTGCTCTCAGCCCCAGT  
AGCAGCCTGCTCTCTCCAGGAGGGCTTCCACTAACAAGTAGGGCATTGC  
TGGAGGGCCAGGCAGACACTAGCTTAGGAAATCCACCAACCCTGGAATG

FIG. 4 (52 of 61)

106/118

CTAGTCCCTTCTCTGAAGGCTCAGAAGAUTGACTTTAGAGTCTAGAAAAT  
ATTGGTCTTGGGAACAGATTTTGTAGTGCAAAGAGATGGACTTCAGATGG  
CCAGATGCACTGCTTCTTTAGGGAATTCTGTGAAAGCTCCCTGCATTTAT  
CTTAATACAGGCAGCAGATTTTCATGAGTACCCCCGAGGGATGGCCCCAGG  
TCCTCCAGCCTGTGAGCATCCTTCTGTCTTCAGCAGCACCACAGTATCT  
TTATATGTCTTTGGATACCTACGTTTCTGCCAGACATCTCTTGCTCTGAT  
GTTCTGGCTGCCAAATTCTCTGTCAAGCGCTCCAATTTTTTGTGTCTCTT  
TGATTTACCCCAACATGACAAAGGCAGTTGTGCTTCATGTATTCAGGGAT  
ACTGCCAAACCACAAACAGGTTAAATCAAATAGCAGATATCCCTGTTCC  
TAAAGACCCATCAGCTCTACCCACCTGCTCCTGCTCACCGTCCCTTATTGT  
TGAGTCTTGAAGCCCTTCTTGTCAATTTTTATTTTTTGCATGAACAATTT  
AGTTCCCTTTGTCTCACTCCTAAACCTTTCTCAAAGGATTGGATTTGTAC  
ACAAACTGCCTATCTCTGCAATCTTAGAAGTGATATGATTCTGAACAAAT  
CACTTAACTTTTGATTTTTTATTGGTAAGATGGGAATACCAATTTTTGCT  
CCACTTCTGTCTATGTTGGCCTGGGCTGATGTTGAAAGCTCTCGGTCAA  
CTGAGATAGGGTGTGAGAATTTATATATATAAATATATCTCTCCAACC  
CCTCCCAATGAAGCAAGTCACGTGAGTCAATCCTACCCTAAGATATTAGG  
GATTGAGCCTCCTGGGACATTTGGTGGCTTAGGTTTCATGAAAAGAGGT  
TGCAGAGCAACTGCTTTTTTGTAGGCAAAGATTAGGCTACTGCAGAGACT  
CAGCAAACCTTCTATAGAAGGTGTGAGATGGTAAGTATTTTAGGTTTGTCT  
TGCCAGATGATCTCTCAACTAGTTAACCATGCTATTGTAGCCTCGAAGCA  
GCCAGAGACAATATGTAAACAAGAGCATGGCTGTGTTTCAATAAACTTT  
ATTTAAAAAACAGTCAGGGACCGGATTTGGCCAAAGGCCATAGTGTGCC  
AGCCCCAAGACTAGAGCAATGCACCTTTTAACTTTTTTATTTTATTTTGT  
AAAATGCCAAGATCCACAAAATGCTATTGCACCCCGTGTGTTAGCACTG  
TGACTCAAGGTTTGGGAAATCTGCTTTGAAGGCGTGATAGACAGGAGAG  
CATGGTCTGGCCCCCTTGGTGCCTTTCTGGTTGCAGCGAGCATTTCAAAT  
ACAGAGCAAGGCCAGTGGTCTGTTTCACTAGAGACATGCAGCAAGGTG  
TCCTGGGGTGAGAAAGATGCCATAACTGGTCCCCTTTCTATCTCCTTAGGT  
CTTGGACTTCAATCCATTTCTGTTGAGTAATAAACTCAACGTTGAAAAT  
GTCCTTTGTGGGGGAGAACTCAGGAGTGAAAATGGGCTCTGAGGACTGGG  
AAAAAGATGAACCCAGTGCTGCTTAGAAGGTAAAGGTTCTTGTAGAAATC  
TACCTCAGGGCCAAAGTGTAATTCCTAGAGCAGAACTTTGCTAGGTGCTG  
TGCACAGACCCAGTTGTTTCTGCTGACTTGACAGTAAGTGAGCTTTCA  
AATTTCCCTGGACAAATAACTAGACAAGAGAAATCTGGAAGAGAAAAGG  
AAGCTTTGCTTCAAGTGTCAGGCACATCAGGTAGTAGATAAAAAGGATCGT  
CCTCACCTACAGATTTGGGGCTTTAGCATCCTGTTTGCCAACTGGATGGT  
TGCATATGCTTCAAAATGCACCTCTTCCCTCCCAACATTTCCCAAGTGGA  
GAGCAAGCCTCCGATGAGAAGGAACCTCTTAAGGCTGGGCTGAACAAATGA  
CCCAGGCACAGGGCATCTGAGTATTCATGAGGAACACATTTGGGTGTTG  
CCCATGGGGGACAATAGGAGGAGGCTTTTGACCCAAATGATTGTCTACTG  
AGGTGTGACGGGAGAGGCCTGTGACATGCCAGAGGCCAAACCCGTGATCC  
AGTTCATCTCTATTCTATGTTTCTGAAGAGGGAAGCTATGATTTAATGTC  
ATTACTATCATGCTGCTCTAGTATTTCTCAGCACATACACAGAAGAGGGA  
ATTAAATGGTCTTGATACCCCTAAATCCTTGGAAAATCCGAATTGCATA  
TGCTAACCTCACTGCGTCTGACTGCAGACCCGGCTGTAAGCCCCCTGGAA  
CCAGGCCCAAGCCTCCCCGCCATGAATTTTGTTCACACAAGTAAGGCCTC  
GGGGTGAGGTGATGGGGGTGGCTGAGGTGCGAGGGTGGGGATGGGGGATG  
GAGCCATTGGGTCTCTTACAGGGTGAGAGAATTGTAGAATGGGGACACC  
TAAGGGTGCTGGATGGGGCTGAAGTCTTTCTTTGTGGAAGCAAATCCCA  
TTAGGAGATAACTCTGGGAAAGATGAGCCCGGGGAGGGGAGGTGATGCT  
CACCTGCTAAGAGGCAAAGGGCAAGGAAGAGTTTGTGCCTGGGAACCTTC  
CAGGTGCCTCTTCTGACCATAGCCAAGAGACTGGAGACACAGACCTCCTC  
CCAGCACTGAGGACAAACAGCCATGGGGCCAGTGGGGGTGCAGGGACACC  
CACACCACTAAGGGCTCAGGGCGGCGCTTCAGAGCCTGAACCTTCCTCT  
CATGCTGCCATTTGAACACCACAACACCCTAATAGGAACTGTTAACATT  
GCCACTGTCAGGTGTGGAACCCGAGACAGACAGTGGAGATTCCTGCCC  
TAGGTGACACAGGTAATAAGTGACAGATGTGGAATTTAAAGGTACTATA  
ACGTCTGCTGCTGACTCAGGCTTAAGGCTCCCATCACCTCCTCTTCTC  
AGGACAGAGTCAGGAGGCCTCAGCCTGAGCCCCAGCTCTAGTGCAGGTTC

ATGTGGGAATACTGAGCCTCACTAGTACATGGCAGAGAGGACCAAAATGG  
GACCAGGTGTGTAAGGGTGCCTGGCACAGTTGGGGGAGGCTGCTGTCGCT  
TCTCCACCGCTGCTGCTGCAGTTACCTTTGATGTTTTAGTTTTGTTGTAG  
TTACACCATTGCTGGCTTTGGATCTGCACTGTGTCCACTCCAGGTGGAAC  
CACGCACACAAGCCTCTCTGTGGGCTGTCTGACTTCTCCTTGTCAGG  
GCTGGGATCTCCTTCAAATCTGGCGGAAGTGGTTCTCCAAGTCTGGTCCT  
CAAACGTCAGCAGCATCAGCGCCTAGAAGTGTTAGGAATACACATTCCCA  
GGCCCCACACAGACCTCCTGCCTCAGAACTCAGGGCGCTGAGGCTCTA  
GGGGCTGCTTTAAACAAGCCTTCAGGTTATCGTGACGCACCTTGAAAGTC  
TGAGAGCTACTGCCCTACAGAAAGTTACTAGTGCCCTAAAGCTGGCGCTG  
GCACTGATGTTACTGCTGCTGTTGGAGTACAACCTCCCTATAGAAAACAA  
CTGCCAGCACCTTAAGACCACTCACACCTTCAGAGTGGCCTTGAGAAAGA  
TTTGGGGTCAAGGATCATGAGCGAGAACCACTTAAGAGGATAGTGAAC  
TAGTCTGCATGTGAGACGCTGAGATCCTATGTCAGGCTGTGATAGGAGGG  
AAACAGAAACCAAAGGAAAGAACAGCTTTAAGAAGCGCTTAAGAGGTACA  
AAGTAAAATGATGGTGCTAGAAAAGTAGCTTCTTAAAAGAGCATTTTCC  
AGTCTCACCCCTGGACTAACTGAATGAGAATCTCAGGAGTGTGAGGCCAG  
GTATCTCATGGTCTTAAATGCCACCCACAGGTGATTCCAGTGTGCACC  
AGGGGTGAGAGTCACAGCCTTAGGCCATGCCACTCAAAGGTGTCTTCAG  
ACCAGCAGCACCCACAGCTCTGGGAGTGCATCAGAAAGACAGAGGCTTGG  
CACCACCCACACCTACTGAACCATAGTTTGCAGGTGATTCTTGCACATT  
AAAGTGTGGGAAATGGAAAGCTTAGAGTTCAGCTAGCTCGGTGACTCTC  
AGTCAACCTGCACCTGCTCCATGAACTCAGACTGCCTGGGATGGGCCCAG  
AAAAGCTCCTGAGGAGATTCTGATGTAAGGCAGGGCTGATAACCATGGAT  
CTCATCTGACCCCATATCACTGGGGAGTTACTTAGGATCTTGCCCTGGGGC  
CAGTCATCTCTCCATAGACACTGAGAGTGTCCACGATGCTTGGGGCACT  
ACAGGGTGGGAGGTGGAGGATCACGGGTGAGTCAGATAGGAAGCCTGCTC  
CTGGGGAGCTTACAGTGCTATAGGGCAGCAAGCCAAGGATGCCAATACCT  
GTGTGCAGGTACCCTGACGAGTGCAGAGCGCTGCAGCACCAGAGAGGAA  
GCTACCCTGTGCAGAGGGGGCTGAGGAGGGCTGCAGGGAGATGACAGGAA  
AGCCGGTGTACAGGAGGAGTCTCCCCACTCTTTGGGCATGAGGAGACC  
AGGAGGACATTCTACAGTGAGAAACCCAGGCAGAGGCCATGTGCTTATGG  
CATGGGAAAAGAATGACACCTTAGACTTATTCTCTACATTAGAATTGCCT  
ACCACAGATACCCATATTATAGCTTCACATAGTGTGGTGGTTACTGTGTT  
TTCATATTGTCACATTTGCCATTTTCCAGCCACCCACCCATTCTTGACAG  
TCACTGGCCAGCCTGGGGGCCCTGTTCTTTATCAAACAAGTGCCTGAG  
CTCTTTGCAGAGGTGAGGGTCACCTGTCCAATCAGAGGCCAGGAGGGAAC  
GTTCCCTTTTAAGACCCTACTCTAGGCAGGCCTGGCCCAAATGAGTTGCT  
AGGAGCCCACGCCCTAAGAACCCTCTGAGCACTGTTGTGGCTGGTCCTGC  
TGCTAGAAGTTGTTCTCCAGGGCCAGGTGCAAGATTTGTGGCTTTTCAA  
AGGAGCCACTAAAGCTCCAGCTCAGCCTTGACGGTGTGGGCTCCTGGG  
GGCTTCTGCTCCAAACCCTCCCAACTCTTCCATCACCGCTCCCTTAGCC  
TGGCCAGTGCAGGGATCTGTTCCACTCTAGGCACTGCTGAGGGAATGATG  
CCTCCAGTCAGAGGGTGCAAAAAGAGAGTTAAGAAAAACAATGATTATA  
AAAAGTCCTTTTTATACGCCAGACATTTTCTTGCTCAGGCTAAGTGCTA  
CTTATTTAGTAAGCATTTTAGTTCTCATAACTCCTCTCTCAAGTAGGTG  
CTGCTATTACTTTTCATTTACAGATGAGGACATTGAGGTTTGGAGAGACT  
TAGTAACCTGTCTCTGTCTACAGCAGAGCTGGGATTTGAATCTATCTG  
TCCAAATCTGGAACCCATTTGCTTGACAGAAAGCTTAATTGCTTGTCCT  
AGCAAGATAGAAAGCCTGGGAGTGGAAGAAATATTAGTGGCTGTGATGT  
CTGAGCCACAGGCAGGGTGGAGAGCTAGGGCTGGGGCCCTTGACGTGG  
GGAAGAAAGGGCTGAGTCTTCCATTTTCAATGTGAAGTGTGATATCTGG  
TGATATTGATCTAGGTCCAAAGGTGAAGAACTTAAACCCGAAGAAATTCA  
GCATTATGACCAGGATCACAAGTACTGGTCTGGACTCTGGGAATCTC  
ATAGCAGTTCCAGATAAAAACCTACATACGCCAGGTGACTCTCAGTTTTG  
GCTGTGTTTTCTGCCTCCACCTAGCAGGGGTAAGGCCCTCCTGCTAGGTGG  
GCTCAACTCCATGCTATACCATGCCCATCTCCAGCAGGTGGTGAAGCG  
AGGAGGAGAGGGCCCCAGGGACTAGGGCATCAGATGAAGGGTCTCTAGCAA  
TGACCAGATCTGAAAGTAGTCTTTCTGGAAGGGCTGGAGAAAAAGAAGGA  
GGCAGACACTTAGACTGGAAGAAGAGGAGGCTTAAACCGGTGTGATGGAG

FIG. 4 (54 of 61)

108/118

GGAGAAGTGGACCACAGAAGTCAAGGGAGAGGGACTGTGCATCAGGCCTGA  
AACCCACAGCAGACAGGAGAGACCTTTCCCTGCTCTCAGAACCCACACATG  
TTCTGACTGTCTTTTTCCAGAGATCTTCTTTGCATTAGCCTCATCCTTGA  
GCTCAGCCTCTGCGGAGAAAGGAAGTCCGATTCTCCTGGGGGTCTCTAAA  
GGGGAGTTTTGTCTCTACTGTGACAAGGATAAAGGACAAAGTCATCCATC  
CCTTCAGCTGAAGGTGAGAGTTCTAGCTCAGTTTCTGGGCCTTTGGCTA  
CCCCAAAGTAAAAGGCCAAGATCCTCAATGCCTCTCGCTTTCTGCAAT  
TCTTATCTTGGCCAATATAACAGGGACATCCACCTTTCTGGAAGCACCAG  
GCAGAAGAGCCCCATAAATTCTTCTCTGGTTCTTGGCCCTTCTAGGGAA  
GGAGGAGAGACTCCTCACAGCGGGGAGACAGCAAGGAGCTGAGCACCTGT  
TCTCCTCTCCTGGGCTCACTGGTCTTGGCCCTGGGCGGTGGCGGTCCCC  
TCCTGCTGTGGCCCTCCATGTGGCAAGCAACACAATTGGGCCAGGACCCT  
GGCGTGCTGCTGTAGGGTAGGAGGGTGTGAGGGAGCACTCGGAGGGCAGT  
GTGTCTGCCCTGCAAATTTAGTCTGGATGGAGCATCCTTTCACTTGAGG  
GGAGAAATCTTAGGAAGCTGAATTAGATACAGATCTAAGCCATATTCTCT  
AATTTTAAAACTATAGAGCTGAGATTTTGGTATCCATCTGACTCTTACG  
TCTCTCTCTCTCTCTCTCTCTCTCAGTTTATTTTAACTCTGGGGGACA  
AGAAGGCCTGGAAGAGAGGGCATGATTGCTTATCATCCCTTAAATACCAG  
TACCAAGGCTGACACGTCATCTTTCCCAAGGACCATCTGCCTTCTCTCTT  
TTCTCTCTCTCCTGTGTAAAGGCCCTGGAGGATGAGCACATGTGCTGTGT  
TTCTCTCTCTCTCAAAGCCTGTGCTATCTAATTAATCCCTTTTACCTCACA  
GAAGGAGAACTGATGAAGCTGGCTGCCCCAAAAGGAATCAGCACGCCGGC  
CCTTCATCTTTTATAGGGCTCAGGTGGGCTCCTGGAACATGCTGGAGTCG  
GCGGCTCACCCCGGATGGTTCATCTGCACCTCCTGCAATTGTAATGAGCC  
TGTTGGGGTGACAGATAAATTTGAGAACAGGAAACACATTGAATTTTCAT  
TTCAACCAGTTTGCAAAGCTGAAATGAGCCCCAGTGAGGTGAGCGATTAG  
GAACTGCCCCATTGAACGCCTTCTCGCTAATTTGAACTAATTGTATAA  
AAACACCAAACTGCTCACTAACTTTCTGTCATTGGGTTTCATTTCTCA  
TTCATGCTTTTAAAGGATTTGTGTTTTTAGGATATAGCAAGAAGCTTGTTTA  
ATTACAAAGTTTGGGTTGGAAAGAGACCGGCTTCTGCTTGTGTACTGCT  
ACCCTGAACCATCAGACATGCATGTGTGTGTCATATGCTATGATGTGGCC  
AGTCTGAGTGCAATACTTGCAGCGGAAGGAGCAGCTGGGTGCATGCTGT  
GCTCTAGAATTAGTCTTTCTACTGGGTTTGGTAGATTCTGAGGGCATT  
GATCCTGGGGCAGAAAGTGGCTGAGTCTGTGTCTAGGGTACAGTGTGCAAG  
AAAGAAATGTAACAGCAAGTCACAATCCAGCCAAGTGATAGTGGAAGAGG  
GGTAGTTAGGTCCAGATAAGGAGCAGGGTGACTTGACCTGTGGGAAAGG  
CACAGAGACAAGGAATCTGGGTCAGATGACAGCCAGGAGACCAGGTGAGG  
GAGGAGCCAGGTACTGTCTGGGAGGCTTGTCAACAAGGGCATGGTCTTAT  
CACTAAGCAGGGCTCAGATCCTCATAATGGGGGAGTGGAAGGCTGGCCGA  
ACAGAAATCAGGGCCTGGAACAGAGTGAGGGGGTGGAGACAGGAGACTG  
AGGCTTGGAAATTAGTTTATTAGTTTTAGCTCTTCAGTTACAAGCAATAA  
TAATAGCTTCTAGCTTATTTAAGCAACAAGTATACTACAAAAGGAGCTTT  
CTAGAAGGATATTGGGTATATTCAATTTCTTACTGCTGCTGTAACAAATTA  
CCACCAACTTAGTGGTTTAAACAATGCAATGTATTATCTTGCAAGTTATGG  
AGGTGAGTCTGGAATGTGTCTCACTGGGCCAAAATCAAAGTATCAGCAGG  
ATAGCATTGCTTTGGGAGGCTCTAGGGGAGAGTCAATTCCTTGCCCTTTT  
CCAGCTTCAGAGGCCACCTGCATTCCTTGGCTAGTGGCCCACTCCCATC  
TTCGCTGCTTGGGTTTTTCTCACACTGCTTTGCTCTGACCCTCCTGCCTT  
CCTCTTTCACATATAAGAACGCTTGCAATTTACATCGGGCTCACGTCAAT  
ATCCAGGATACTCTCCCGTCTCAAAGAGGCTTAACTTTAATCACAGATGC  
AAAGTCCCTTTTGTATGTGATGTAACATATACACAGGGTCTGGGGATTA  
GAATGTGGACATTTTGGGGTGCCATTATTCTGCCTATCATGTGAAGTAA  
CTTTCAAATGGAAGACATGCTGAAGAAAAAGTCAGGGATTTCTGGCAG  
GCCAGAAATGACAGAAGGCAGAAAACGTTGGTCCCATCACTCAGATGGGT  
AAGAGCCAATCATGCTTTTGTGAGTTAGCAAAAGATTGAGATTCCAAGC  
AAAGCATGCAACTGCCCTAGTTTGGGTGATGTGTGACTCCTTGGTCAGT  
GAAGGCGAGCACACCTTGATCAATACTCCCTCCAAGACTGTATCCAACGA  
GGCGTGTGAGTGTTCCTCAAAGCAGAGCTAGAGAGCTAATCCCAGGAGAGA  
GGCGTGTGGGTGGTGGGCAGGAAGACAAAGCTCAGCCGTAAAGGAGTAGT  
AGGGACAGCACCTAGGCATGGAGGCTCAAGTGAGATGATACCCATGGGA

AAAGCTCTGATAAGGTCAGCTCCTTCTGTTTCTGATCCTGATGGTGATGG  
TGATCAACACCAGCCCAGTGACAAAAAGTACATAGTATATTTAGTAGAT  
GTTTCCCACACAGAGAAATGGTAAATATTCAAGGCGAGGAATACTCCAAA  
CATCCTACCTTGATCATTACACATTCCGTGCATGTAATGAGTACTTGCAT  
GTATGCCATAAATATGTGAAATATTATGTATCACTATATAAAAGAAAAAA  
AAATGTGGCCAGGTGACATCCATATTTTGGAGAGGAAGGCATGTCTTCTT  
CATAATATCACAAACTATTTTCACAACAAAGACACAGCTGTTCAAATTA  
GTCTCTGAGCCGGGCTGTCTCATGGCAGTGAGGACTCTGGTTCCTTAC  
AGACTAGCAGAAAGGAGATGGGGCTTACTGACCATGGCCTTGAGGAGGCT  
GAACATGCAGGCCAAATGGAGACACAGACAGCCTGGGCTTGGTCTTGCTC  
CATCCCCCTTCCAACCTGATGAGATATAGTGAGTCACTATGACGTGGGTCA  
CTCATGCTTCCTGTGAGGCTCCACCAAGACAGCAAGTGCATCAACACCTT  
ACGGAAGCACAAGGCCCTGTTTGTGTTGACTTCATGAAAGGCATGGTTG  
TGGTGATCGCATTGAGTAGGCTTTTGGGTGAGAGGTGAAAAACCCCACT  
ATCATGCATTGCAGCCCTCTGGTGGAACTGTGCTTCAGGCTCTAAATTT  
CAGGCTCTAGACTGACTCCAGGATGAGTATTTGGAAGCTGAAGTCAATCT  
GTGGTCTCTTCTCCTGTAGAGCAGGAGTCAGCACTTTTCATAGAGTGCCA  
GATTCTATATATCCTGCCACATGCTCTGTTGTTACAGAACAAAGAAGGCC  
ATAGACAGCATGGCTGTGTTGGCAAATACACAAAACAGGCAATAAGCTGT  
ATTGGCCCTTTAGGCTGCAAGTTTGCCAAACCCCTGCACTAACACAGAGCTT  
AAAGGTGGTGGTGGTGTGCTGGAGCTAGCTTATATCAGCTTGCAATAGCC  
AATTGCTAACATCTCTTCCAACTCTGTGTCTGTGCCTTGATGTTGATAG  
TTTGAAATTGGCTACCCCATTTAATGCTGCAATCTTTTCTACCCCAAGCA  
CTACTGACTCCCTTTGCCCTGTCTTATTTTTCTCACTCTAACATGCTGT  
ATAGTTTTCTTCTTACATTTATTGTTTGTGCTTCCACTAGCATGTATGT  
CCCACAAGTTCTTTGCTCTGTGATGTATCCCAAGAACCCACTGCAGTGCT  
TGGCAGTTGTAGGAACTCCATAAGATTTTATAAATGAAGAAAGGAAGAA  
AAAAGAGAGGGAGGGAAAAAGGAAAGGAAGCCTTCTATTTAAATGATGGC  
CTTCTCCATATTTCTATAGTAATATGACTTCCCTTGCAAAGGGGGATGCA  
TTTTGGAAAATGTGTATAAATAAACTCAGGTGGTTTTGAATTTCATTTTCT  
CTAACTGTAATTGTAATCATTGGTCTTTATGTTTAGTGAAAAAGTTTTGG  
CCCTTATGCCTCACACCTGAGAATCCCAAAGTATTGGTTTGTAGAGCTC  
CCATAGAGAACCATAAACTGGGTGGCTTAAACAACAGAAATGTATCGTC  
TCCTGGTTTCAGGAGGCCAAAGTCTGAACCTCAGGTGTTGGTTCACTCTGA  
GAGCTCTGAGAGAGAATCTGTTCCAGGCTTCCCTTCAGTTTGTGGTAGCT  
CCAGGGTTCTTGGCTGGTGGCAGCAAACTCCAGTCTCTGCCCCCATCT  
TCACATGACTGTCTTCTCTCTGTGTTTCTGTGTCCAGATTGTCCTATAAG  
GACAGAGTCATCTGAATTAGGGCTCACTCGAATGACTTCATCTTAAGTT  
GAACTGTATCTGTAAAGACCTTATTTCCAAGTAAGGTACATTACAGCT  
ACTGGGGGATAGGACCTCAACATATCTTTTTGGGGGACATAATTCAACTC  
ATAATACCCAACATGATAACTGTTTATCCCATGAAATTTAATGTCTCTCA  
AAAGGTGATCTCAGGGCATTTAATCTGTGACAGAACTCCCATAGGAAAC  
ATTCCAACCAGAAGCTCCTTTACAGCTGGTCACTCCTCCTACCCCATCC  
GAGGTCTTGGGGCAGGGTGAGGCAGGTGGGGACAAGAAGGCTGTCTC  
GGGTGTAGAAAGAGAAGACCTTATTCACCCGGCACTCTGTTTCATGAATG  
AGCTATCCAGCATAGGATATAATAAATCGCTTTAGGAGTGGTAGACTCCA  
AACATTTTTTGGTCCCAGTTATCCTAATCAATTAAACAAACTCTAGAAC  
CCATCTTGAAGTGCAGGCATTGGGACATTATGAAACTTACACAGAATTCA  
AAAATTTACAAGGGCTAAATAAAACAGGGTCTGACATCTAATATTTTCTT  
CCCACATTCCCATGCACTGTCTGGCTCAACCATCCCCAACCCCTCACTCTC  
ATCCTGGTGGACACATGCCTAGTGATGTGATCAGCTGGTTCACAGGGGGC  
TGGTGATGGTGGATATACAGCTTTTGCCAATTTCCATGGCATAACTACTC  
CAAATATGGCCAATTTCAAACATACCAACATGAAGGCACAGACACAGAGTT  
TGGAAGAGATGTTAGCAATTGGCTATTGCAAGCTGATATAAGCTAGCTCC  
AGCACAGCACCACCGCTACCTTTAAGCTCCTTGTGTTAGTGCAAGGGTTG  
GCAAACGACAGCCTAAAGGCCAAATACAGCTTACTGCCTGTTTTGTGTAT  
TTGCCAACACAGCCATGCTGTCTATGGCCTTCTTTGTTCTGTAAACAACAG  
AGCATGTGGCAGGATATATAGAATCTGGCAGTCTTTAATAAGTGCTGACT  
CCTGCTCTACAGGAGAACACAGATTGTCTTCAGCTTCCAAACATTCATCT  
CTGAGTCAGTCTAGAGCCTGAAATTTAGACTGAAGCACAGTTTCCACCAG

FIG. 4 (56 of 61)

110/118

AGGGCTGCAATGCATGAAGTTGGGGTTTTACCTCTCACCCAAAAGCCT  
ACTCAATTTTTTACTGCAAAAACATGTTATCATCATTATTTTTTACTTAG  
CCCACCTTTCTTGGCAATTTTTCCATAGGAAAATGCATTCTAAATTTCAA  
CTAATCAGGGGACTTGGAGCCTCTGGACACCCCTTGTTCCTTGCCACA  
GTCCCTTGAGAAAGGTGCCTTATCAGAGCGGCTCCATGCAGGGGCTCAGG  
ACAGGATCAGATGTCAGTTGCACCAAGGGGGCAGGGACAGATCCTCTCTG  
CTEACCATGCAGAAGGGACTGTTTCAGTGCACCGTCATGGTCCTGGTGATT  
TCTGGTCCATAAGGGAATTTTACATGCATCGGGTGATTGTCACATCAGC  
ACAACACTGTGAGGAAGGCAGAGTGAGAATTTGTGTGCCCATTTTATAGG  
TGAGAAAACAGATGCAGAGACATTAAGTAACTTCACCACAGTCATGCGGG  
TTTTAAGTGGCAGACTTTTCAGGTGTGTGACTCCTAGTCCAGAGTTCTTT  
GCACTGCCCCCTGAGGTGCTAAACTCTACTGTGCTTTAAGACTCACTTGG  
GGAGCTTCTTAAAAAGAGAGATTGCACAACCTGAGATTCTTGTTTAACTG  
TTTTGGGATGTAGCTCAGGGATCTAGCTGCCTTAAAAAATACTCCCA  
AGTAATTCTGATGCAAGCGGTTCTTTTTTGTCCACCTTGAAGAAACT  
GCCTCCTCCCCATACATTTTATTAGAAAATGGTAACATGTTTTTCAGCCT  
GAGAGCCATTTCTGGGTGACCGGACGTCCGGCAGCCCGCTGTACTAGCTTT  
CAGTCTAGGCTTAAACACACATGATAGGAGATGTCCTACTCCAGATGATA  
TGAGTCTGAACCATGGAAAAATTCATTGTGTGGCACATCTGGTGGGTGT  
GCACTGTCCCAGCAGTGAGGCACCCAGTGAAGACAGCAGCTGGGAGAGG  
CTTAGTTACATGAGTGGGACAGTGTGGGCTAGACTGCTGAGCCCTCTGC  
AGTTTTACTCTGTGTGAGGCAATGAGGGTGAAAGGCTGATCAGACCCACGT  
GCAGACCATACCCTCCAGGGAGACAGATATCAGTCAGGACAACCCCAAGT  
GTAGCTGGAGAAGCAGTGCCAGGTATGACCGGATGTGTATCCAACCAGG  
AAATCTGCATATAAATATAAGAGGAGAAAATGAACAGATGTTGCTCTTAT  
ATGTAGATATTTATGAAGAGCATATAATTTTGTGTGTGTGTTTAAAGAA  
GTTTATAAGTATGCCTTAAAAATGTATAGTATATACTGTAGGTATTTTT  
CCATTAGATATTTTGTGTTTTTCATACTTATCCACATTGACATTGTAGCAAC  
AGTATAATATAACAACCTCCTCTACAAAAGCAGAAGGAAGTGAAGCTTTG  
GAAGGAAGCAGCCAGTGAGCTTGCCCTTTTCAGGTGGGTGCAGTGAGCAG  
GAGTCAGTGAGGTGAGATCCTTTGAGAGGAGGCAATCATTAAACCAGGAA  
ATCTGCACTGCATCCTGGCCACACCTAACCTTGGACAATGGTGCTTGGA  
GCGCCTTCCAGCTCTTAAGGCTTGCGATTTCTTCTCTCACTCTTCAACC  
ACGATGATTAAATCTTCTCCTACAGAGTTGGACAATAAAGCCTTGAGTTC  
CTGCCCTCCCCTGGTGTGATCACGAGGCATAGACATGGCCAGGAACATGTA  
GGTGTCTTTGAAAGCTGAACAAGTTAGTAAATTTCAAACCTCATTTCAAC  
CACCAGTAAATGGGAATAATAATAAACCTATTTTACATAGGGTTGACAA  
GAGGAGTAAAGAGGGATTCAATGAAAGTTCGTTATTATCATTGTTAGTAG  
CAGTGTGATAATATCAACTGAAAGTTCATTATCATTATTAGTAGCAGTA  
TTGATAACCCCTTTTCTGTGCTTCTCACTGGTGGGCCAGGCCATCAG  
CAATGCCAGGGGTGCATGGATCTCTGCTGCATCGGGCACCAGCTGTGTC  
AATGGTGAGAACAGTACAAGGGTGGGCAGGGCAAGGCAGGAAGCAGCCAG  
GAGCAGCAGCTTCATGGGGTGAAGATGTCAGGAGCTTAGGGACAGTCAGA  
GCGGGTGTGCCTCCTCTTGTGGAGCCTTTCTGCGTGGGTAGGAACTGCTG  
CAGCTGTGGCCATGGATTACCTGAATATGGGTGGAATTAGGCATTGAGC  
TGGGTAGCTGTGCCTAGAAGGAGGAACTCTAAACTGAGAACTTGTCCCT  
ATTGCCACCTCTGATAGGCAGATGATCCATCCATCAGTGGCTGAGCTGAG  
GTGTGCATGGGGATGGGTAAAGAGCCACACACAGGGCTGATGACTGAGTC  
TATTTAGAACAAATAGATGTAAAATCTGATAATGTAAAATGTGATAGATTA  
TTTTGTCAATTAGAAATGGTACCATATAATTATATATATACATAAACATG  
TATACATATAACACATATAACATGTGTGTATAAACACACACAGTATTGTC  
CCCTACTCATTCCATAAACCTGATGCCTTTAGCTGGGATTCCCAGCTTTC  
ACTCTCCTCTCTGTCTGTCTGTCTATATCCTCCCCATCCTGTAATTCT  
GGCTTATATGCCACTTCCCTCCCTAAAGCCCTCCCTCAATCCCTTGCTGGA  
AGTGACATTTTCTCTTTGAGCTGCCCCCTGCTTGTGCTTTGGTGAGGTCA  
GCTGTATTGCAGTACCTTGTATTGTGGTTGTACATCATCGTATAGAATT  
AATTTCTGACACATTCCGTATTTTTCAAAGGGCCTAGTGTGGGGCTTTAA  
CAGTAACTACGCCACCACGCCAGTTAATTTTTGTATTTTTGGTGAGGA  
CAAGGTTTACCATGTTGGCCGGGCTGGTTTCGAACCTGACTTCAGGT  
GATCTGTCTGCCTCAGCCTCCTGGAGTGCTAGGATTGCAGGCATGAGCCA

CTGCACCCAGCCACCTATCAAAATTTTAAGTGCCATTTTTATTATTTTATT  
TTTTGTAGAAATGGACAAGCTGATCGCAAAATTCACATGGAATTGCAGGA  
GGTTCCAAATAGCCAAACAATCTTGAAAAAGAAGCAAAAGTTGGAGGA  
TTTACACTTTCCAGTTTCAAGACTTAGCTCTTAGCTACAAAGCTACAGTA  
ATCAGAACACTATGGTCCTGGCATAAGTGATGCTGGACAGGTGAGCCCCA  
AAGTGGGACTTAACCTGTGAAGTTCTTGGCCTTGCCCAGGAAGGAATTC  
AAGGGCAAGCCAATGGGACAAGAAAACAGCTTTATTGAAGGGGCAGTATT  
ACAGCTCCAGCCCTGTTACAGCTCCAGCCCTGTTACAACCTCTGACTACTC  
CTGCACAGAAGGGCTACCCTGTAGGCAGAGAGTAGCAACTCAGGGCAGTT  
TTGCAGTCATTTATATCCACTTTTAACACATGCAGATTAAGGGACAATTT  
ATGCAGAAATTTCTACGGAATTGGTAATAACTTTTGGGTGATGGAGTCAT  
CATGGAAGGGGGGCGGGGAACCTCCCTGGTGTGCCATGATGACGGTAAAC  
TGATATGGCGAACTGGTGGGTATGTCACATGAAAAGCTCCTTCCACCCCA  
GCCCTGTTTCAATTAGTCCTCGGTTTGGTCCAGTGTCGAAGTCCTGCCTC  
CAGAGTCAAGTCCCACCCCTACCTCTTAAGGAGAGATGTAAATACATGG  
AATAGAATTGAGAGTCCAGAAATAATCTCATACTATGATCAATTGAT  
TTTCAGCAAGGTGCCAAGACCATTCAATGAGGGAAAGAATCATATTTT  
TTCAACAAATGGTGTGGATAACCACATGTGAAAGAATGCAACTGGGCC  
TTATCTCACACCATATACAGAAATTAACCAAAATGGCTCAAACACTTAC  
ATGTAAGAGCTAAAACTATAATATTCTTAGAAGAAAACAGGGATATATCT  
TTATGACCTTGGATTTGCTGGCTGATTCTTAAATGACACTGAAAGCACA  
GCAACAAAAGAAAAAAAATAGGTAAATTGGACCTCATCAAAATTTAAAA  
CTTTTATGCTGGGTGCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGG  
CAGGAGGATCTCTTGAGCCCAAGAAGCTGAGGCTACAGTGAGCCGAATT  
GTGCCACTGCATCCAGCCTGGGTGACAGAGCAAGACCCTGTCTCGAATA  
AATAAATAAACAAATATATAATTATAGATCTCTGGATCTTGCTTCCGAG  
ACTGACTCAACTAACTGGTCTGGGTGGGAGCCCAGCCATTTGTATTTTT  
GAAACTCTCAAATGATTTTACTGTGCAGCCAAGTTGAGAATCACTGT  
ATCATAGGGTTGGACTCCTAACTGGAACAGTTTGCACCATCAGGTGTG  
CAGCATTCTGATAATAGTTAAGCTTTCTCTTAGATTTTCTGATATTAGA  
TGAGTCATGTTTACAAGTTTTTACCAAGAGACAACTATCTTTCTGCCCT  
TACTTTCTCTTATACTATTCTAATCCCAGAACCCTTTGGAACCTCCAC  
TGAGAGATGAATCTAGAAAGTGACTCTCTTGGCTACAACAGAGAGTAATG  
TTGGCCTGTTTGTGCCAGATCCAGTTGGTGCTGGTGGTGGGACAGCACCT  
CCCTGAAATCCCCTCCTCTCCCGTCAGATTAGTCCCCCATTTGCATCAC  
GTACAATCATCACTATGGGTTTCTATTACCTTGCTAGGGCATTGAGGT  
ACCATATATACCAACTATTAGTTTTGAGCCATGGTTCCCAAGTGTGGAC  
TGTAGGGCACCTCAGCACACTCAGAGGTGTGATGGGATATTTAAATATT  
CTGAAGAAAACACAGTGACATCTGTGAGGCCCGTGAAAACCGTTGGCATT  
AAATTGTCTCAACCCAATTGCTTAAGAAGCAGAACTGGCCAGGCACGGTG  
GCTCACATCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCAC  
GAGGTCAGGAGTTGAGACCAGCCTGACCAACATAGTGAAACCCCGTCTC  
TACTAAAAATATAAAAATTAGCCATGCATGGTGGCATGCACCTGTAACCC  
CAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAAGCGG  
AGGTTGTAGTGAGCCAAAATCGTGCCACTGCACTCCAGCTTGGGTGATAG  
TGAGACTACATCTCAAAAAAAAAAATGAGAGAGAGAGAGAAGCAGA  
ACCATCAGGTGTTTCTTTTGGCTTAAAGTACTCTGTGAAGAAATTCCTGG  
GACACGAAGGATACCATGAACTGAGAGATTTTGGGAACCTCTGCTTTAGA  
AGCTGGAGGTAGCATTCTTGGGCACAGTACTGCCTTGGGATCAGCAAT  
CCTTTTGTAGGTGCATTTAGGTGTGGCAAGACAGCTCTTAGAGTGGGACC  
GGGATGTGCTTGGAGACAGAGGGAAGTAGATTGAGCTGCCCGATAAAGAC  
ATGCCAGCCTGGCAGAGTGTAGTGACTCATGTCTGTAATCCTAGTGCTTT  
GGGAGGCTGAAGTGGGAGGATTGCTTGAAGGCCAGGGGTTTGAATCAGCC  
TGGGAAACAACAAGACCTCTACAAAAAAGAAAAAATAACCA  
CATGTGTTGGCATGCACCTGTAGTCCAGCTACCTGGCAGGCTGAGGTAG  
GAGGATCACTTGAAGCCAGGAAGGTAAGGATACATTGAGCCATGACTGTG  
CCACTGCACTCTAGCCTGGGTGACAGAAAGAGACTCTGTCTCAGAAATAA  
ATTAATAAATAAATAAATATATAGTGGCCATGACATCCCTAGAAAGACA  
AGGTCCTGGGAATAGGTAGAAGCCAAGGGAAATGAGAAATGAGAGGGGGC  
CCTGGAGCTGGAAGTGGGGGAGCAGGATGGCCTCTGAGAAGTCTCTGATA

FIG. 4 (58 of 61)

112/118

FIG. 4 (59 of 61)

TGGTCCCGTGAAGCCTCACACATGGTACACAAAGGCTGTCTTGAAAAGA  
AACGTAAGTGTGTTTTTGGTTTAATAAAATTGATTATAAATGGATAATG  
CAAAACATTTTAAAGAATTTTACTAGCTTACATTAGCAGATTTGGATCCA  
GTGATTGTTACATTCTGGTACTGAGCCCCTGAATTACTTCTTTGAGTAAG  
GCATTATACCAAAGCTATTGATAGTTGGGCTTATAGGGTGTATGTTTGAA  
GAACTACTAATGTCAAACCAATATTTACGGTTCGACAAGAGGACATCAG  
AACTGGTAATCCTTATTACCATGACTGGCTGGACAGAATACTCAATGTAA  
TGGGATTTCTGCAAATAAAGACGGGGAAGATGTAAAAAAGATGCCTGAA  
CATTCAACATTAATGAAAGATTTCAGAAGAAATATGTATACTAACTGCAG  
CCTTATCAAGTATATGGAAAAACACAAAGTTAAACCAGATAGTAAAGCAT  
TCCACTTGCTTCAGAAGTTTCTTACTATGGACCCAATAAAGTGAATTACC  
TGAGAACGGGGTCCCTGTTTCTCGAAGACCCACTTCTTACATCAGACGT  
TTTCAACAGTTGTCAAATCCCCTACCCAAAATGAGAATTTTAAACAGAAG  
AAGAACCTGATGACAAAGGAGCCAAAAGAACCACCACCGGCAGCAGGGC  
CATAACCACACGAATGGAAGTGGCCACCAGGAATCAAGACAACGGTCAC  
ACACAGGGACCCCCGTTGAAGAAAGTGAGGCTTGTTCTCTCTACCACTAC  
CTCAGGTGGACTTTTACGGCCTCAGACTATCCGCGTTCCAATCCACATG  
CTGCCTATATCCCAACCCTGGACCAAGCACATCCCAGCCGAAGAGCAGTG  
TAGGACTACTCAGTACCTCCCAGCAGGCTCCACAGGACCCACGTGAGACA  
CACGGGTACTGAGCTGCATCGGAATCTTGTCCTGTCAGTGTGTGAATGC  
TGCAGGGCTGACTGTGCAGCTCTCCGTGGGAACCTGGTATGGGCCATGAG  
AATGTACTGTACAACCACACCTGCCAGTAGCCAAGTTCCTTCCACCGCT  
TTTCACAGATCGGGGTAGTGGCTTCCAGTTTGTACCTATTTTGGAGTTAG  
ACCTGAAAAGAAAGCGCTAGCACAGTTTGTGTTGTGGATTTGCTACTTTC  
ATAGTTAACTTGACCTGGCTCAGACTGACCAGTACTTTTTTTTCCGTGAC  
AGTCTATAGCAGTTGAAGCTGAGAATGTGCTAGGGGCAAGCGTTTGTCTT  
CATATGTCATGAATTCCTCCAGTGTAACAACATTATCTGACCAATAGTAC  
ACACACAGACACAAGGTTTAACTGGTACTTGAAAACATACAGTAGGTGTT  
AACTCAGTGAAATAACCAGGACTCAAAGTAAGATTATTTTGGTACACCTT  
TCTTGTTAGTGTCTTATCAGTGAGTTGATTCAATTTTCTACATTAATCAGT  
GTTTTCTGACCAAGAATATTGCTTGGATTTTTCTGAAAGTACAAAAGCC  
ACATAGTTTTTTTTCAGAAAGGTTTCAAACCTCTAAAGATTAATTTCCAA  
GTATAAGTTTTGTTTTTATTTTCAATCTATGACTTGACTGGTATTAAGCT  
GCTATTTGATAGTAATTAGATATATTCTCATTGATATAAACCTGTTTGGT  
TCAGCAAACAACTAAAATGATTGTCACAGACAATGCTTTATTTTTCTGT  
TTGGTGTGCTTGTGGGAAAAAGAAAGAGAGATCAGATTGTTACTGTGTC  
TGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTCTGTACTAAGA  
AAAATTCCTTCTGCTTGTAGATGCTGTTAATCTATATAACCTTACCCCCAA  
CCCTGTGCTCTCTGAAACATGTGCTGTGTCCACTCAGGGTTAAATGGATT  
AAGGGCGGTGCAAGATGTGCTTTGTAAACAGATGCTTGAAGGCAGCATG  
CTCGTAAGAGTCATCACCCTCCCTAATCTCAAGTACCAGGGACACAAA  
CACTGCTGAAGGCCGAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCC  
AAGGTTTCTCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAGGGGAA  
AGACCTGACCGTCCCCCAGCCGACACCCGTAAAGGGTCTGTGCTGAGGA  
GGATTAGTATACGAGGAAGGAACGCCTCTTTCAGTTGAGACAAGAGGAA  
GGCATCTGCTTCTGCCCCGTCCTGGGCAATGGAATGTCTCGGTATAAAA  
CCCCGATTTTATGTTCCATCTACTGAGATAGGGGAAAACCACCTTAGGGCT  
GGAGGTGGGACATGCGGCAGCAATACTGCTCTTTAAGACATTGAGATGTT  
TATGTGTATGCATATCTAAAGCACAGCACTTAATCTTTACCTTGTCTAT  
GTTGCAGAGACCTTTGTTACGTGTTTATCTGCTGACCTTCTCTCCACTA  
TTATCCTATGACCCTGCCACATCCCCCTCTCCGAGAAACACCCAAGAATG  
ATCAATAAATACTAAGGGAACCTCAGAGGCCGCGGGATCCTCCATATACT  
GAACGCTTGTCCCCTGGGCCCCCTTATTTCTTCTCTATACTTGGTCTCT  
GTGTCTTTTTCTTTTCCAAGTCTCTCGTTCCACCTAATGAGAAACACCCA  
CAGGTGTAAAGGGGCAACCCACCCCTTATTGCTGATTTGTGAGCGTGCT  
TTAAGGTGAAAAAAGCATGAATGTTAACTTCCTTAAAAAGGTACAGCATC  
CAATTCAAATATTTTGTCTGATTGTTTAAATGCTAGTTGATGTAGTGCTAT  
TAAAATTTTGTTCACATGGACACAGAGAGGGGAACAACACATACCAGGG  
CCTGTTGCGGGGTGGGGATGAGGGGAGGGAACCTTAGAGGACAGGTGAACA  
GGTGCAGCAGATCACCATGGCCACATATACCTATTTAAACAACTGCAC

GTTCTGCACACGTATCCATTCTTTTTTTTTTTAAGAAATAGAAAAAA  
AATAAAATTTGTTCACTGATTCTTCCATTTTAAACTTGTTTGCATGTG  
GTTTAGGATGCCCTTACTTCAGCAAAGGAGAAGGAATAGGAGGGCCTTAG  
AATTTTGTAGGGAAAAAAACCCTATAACATACATTGTACTGTATCAAACT  
ATTTTACATGAATGACACAAGTATTCTGAATAAAAAAATAATTGAACATT  
GTTAAGAACAAGGTGTCATGTAATTTATTTTTCATAAATAAAAAAATTAT  
AGTGGCTTAGACTGAAAGGAACAGAGAATTTAAAAAATTAAAAAGAAGCC  
TTAGTATATTTTTGTATATAGTTTCCATGTGCCATATTTGCCATAATTGG  
ATGAGAATTTTTTGACCTCTGGCAGGGTGACCCTATATTTTCANTNTATA  
AAGCGTGCATCATAACC

MVI.KCHPPGDSQCAPGVRVTALGHATQVSSDQQIIPQI.WECIRKTEAWIIHPIILI.NIISI.QPGIPCSI.SNKCI.SSI.QRSASA  
EKGSPILI.GVSKGEFCL.YCDKDKGQSIPI.SI.QI.KEKI.MKI.AAQKESARRPFIFYRAQVGSWNMILESAAIIPGWI:ICTSCNCN  
I:IPVGIXNXVDI:DI.I.GKAQKRGTCSE

FIG. 5

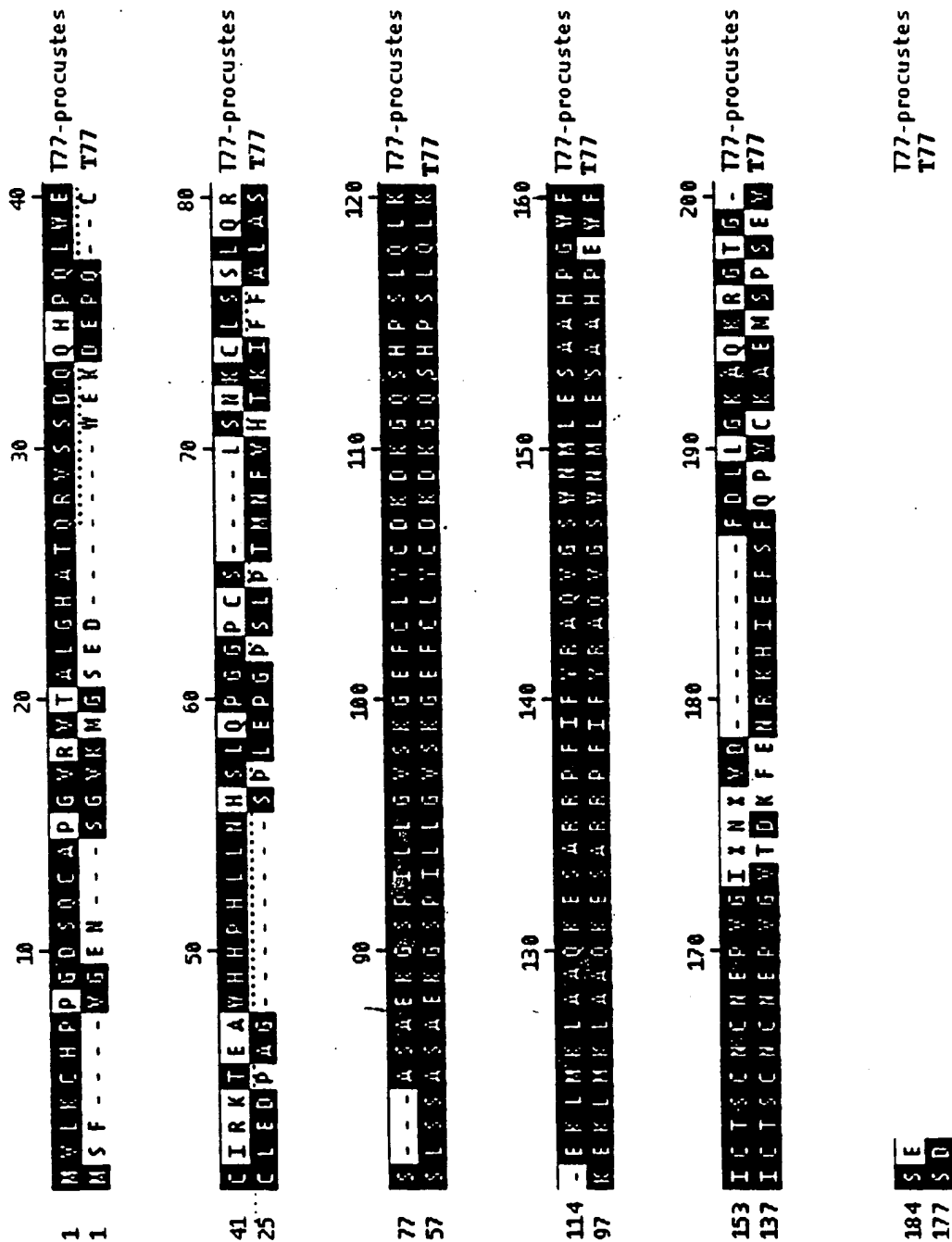


FIG. 6

117/118

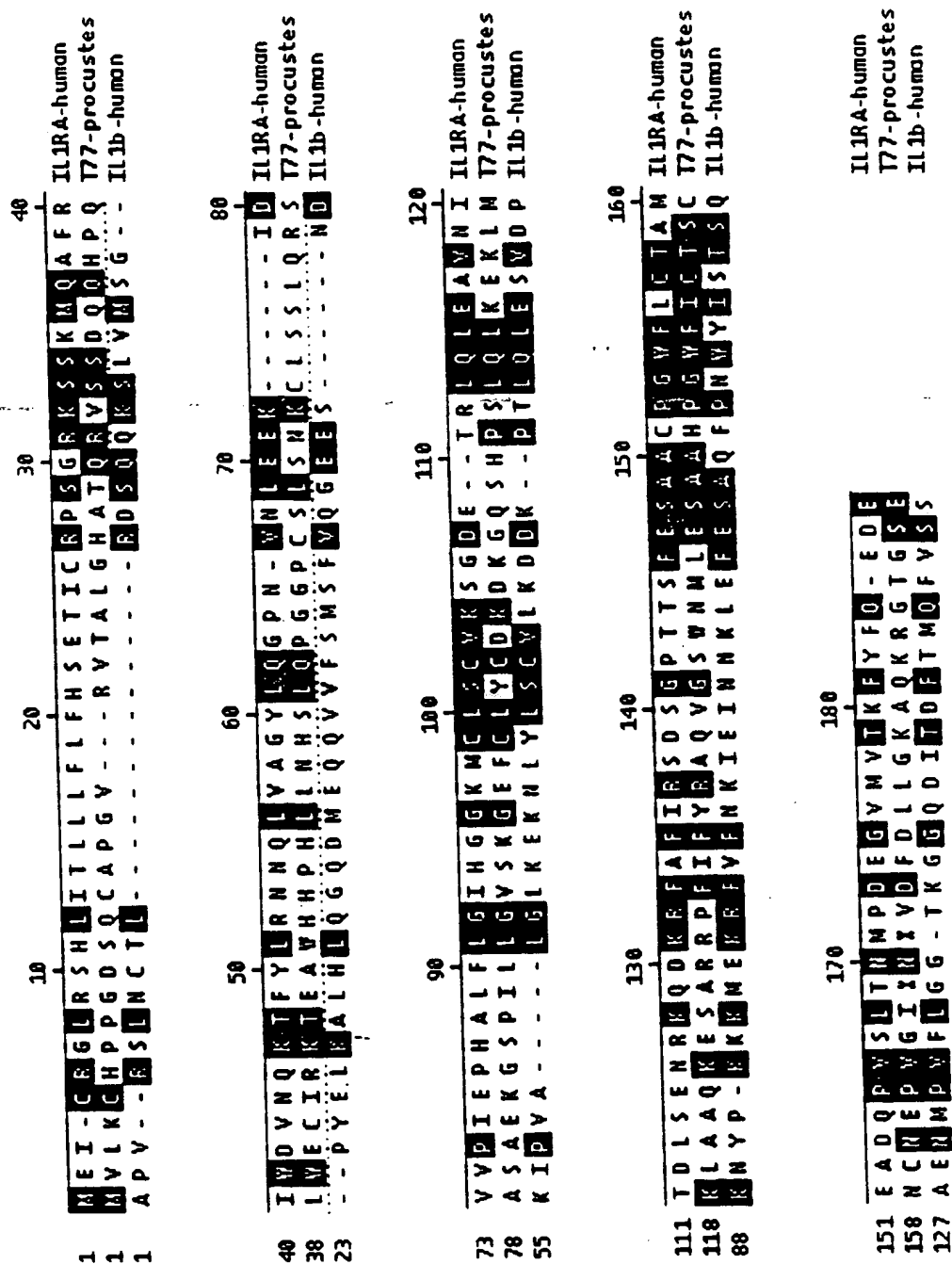


FIG. 7

118/118

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/16102**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C07H 21/02, 21/04, 1/00, 14/00, 17/00; C12Q 1/68; G01N 33/53

US CL : 536/23.1; 530/350, 387.1; 435/6, 7.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 530/350, 387.1; 435/6, 7.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG: MEDLINE, USPATFUL, WPI, BIOSIS. Search terms include author, "TANGO" and protein

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| A         | Database Medline on Dialog, US National Library of Medicine, (Bethesda, MD, USA) AN 09370320. SONNENFELD et al. 'The Drosophila tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development'. Development. November 1997, volume 124, number 22, pages 4571-82, Abstract. | 1-22                  |



Further documents are listed in the continuation of Box C.



See patent family annex.

|    |   |    |  |
|----|---|----|--|
| •  | Special categories of cited documents:  | •T | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| •A | document defining the general state of the art which is not considered to be of particular relevance  |    |  |
| •B | earlier document published on or after the international filing date  | •X | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| •L | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | •Y | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| •O | document referring to an oral disclosure, use, exhibition or other means  |    |  |
| •P | document published prior to the international filing date but later than the priority date claimed  | •A | document member of the same patent family  |

Date of the actual completion of the international search

21 OCTOBER 1998

Date of mailing of the international search report

30 OCT 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

HEATHER BAKALYAR

Telephone No. (703) 308-0196

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**